Investigating the At-Risk Mental State and First Episode Psychosis using Genetic, Cogntive and Multi-modal Neuroimaging Data
A Multivariate Support Vector Machine Study

Pettersson-Yeo, William

Awarding institution:
King's College London

Title: Investigating the At-Risk Mental State and First Episode Psychosis using Genetic, Cogntive and Multi-modal Neuroimaging Data

A Multivariate Support Vector Machine Study

Author: William Pettersson-Yeo

Investigating the At-Risk Mental State and First Episode Psychosis using Genetic, Cognitive and Multi-Modal Neuroimaging Data: a Multivariate Support Vector Machine Study

William Shang-Chin Pettersson-Yeo

BSc (Hons)

King's College London, Institute of Psychiatry

Thesis submitted to King's College London in partial fulfilment for the degree of Doctor of Philosophy (PhD)

September 2012

ABSTRACT

Numerous studies report significant biological and cognitive alterations in chronic schizophrenia (ChSz) patients relative to healthy controls (HCs). More recently, similar, albeit less severe, changes have been reported in subjects with a recent first episode of psychosis (FEP), and those at clinical high-risk, referred to as the at-risk mental state (ARMS). The clinical impact of such findings has been limited, however, driven in part by the univariate analyses employed by the majority of studies, which allow inference at the group level only. Support vector machine (SVM) is one alternative multivariate analysis, which, able to provide inference at individual level, has high potential for translation into a clinical setting.

Here, I employed a multimodal approach comprising genetic, structural magnetic resonance imaging (sMRI), diffusion tensor imaging, functional MRI, and cognitive data, in order to investigate the capacity of each modality to distinguish FEP and ARMS subjects from HCs, and each other, both at the group, and the single-subject level, using standard univariate and multivariate SVM analyses respectively. Since the clinical potential of SVM is ultimately governed by its classification accuracy, I also performed an empirical comparison of four integrative methods, proposed to enhance classification through data integration.

Collectively, the results provide relative support to the notion that FEP and the ARMS may be characterised by genetic, neuroanatomical, neurofunctional and cognitive alterations similar to those observed in ChSz, albeit less severe. With respect to neuroanatomy, and neurofunction moreover, they suggest such changes may be both subtle, and spatially diffuse. The achievement of only modest classification accuracies, however, suggest that the modalities investigated here have only limited diagnostic

power with respect to early-stage psychosis, though it remains that they may be able to provide useful information for predicting conversion to psychosis or treatment outcome, a prospect which could be investigated by future studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This work was funded by a studentship from the Medical Research Council, with study costs also provided by a project grant from the Wellcome Trust.

I would like to acknowledge all the people who have provided supervision and guidance at all stages of the work towards my PhD, most notably, my principle supervisor, Dr Andrea Mechelli, for his considered and insightful thoughts, generous allocation of time, and consistently inspiring foresight, and also my second supervisor, Dr Paul Allen, for his constant all-round support, ready-availability, and consistently constructive advice. It has been a privilege to work with them both.

I would also like to thank friends, colleagues and fellow PhD students whose encouragement, assistance and support throughout, have been invaluable. In particular, I would like to thank, Stefania Benetti, Lana Ilankovic-Kambeitz, Graziella Orrù, Dr Diana Prata, Dr Chris Chaddock, and Dr Flavio Dell'Acqua. In addition, I am especially grateful to Dr Andre Marquand, for his help and guidance in navigating the exciting, and often complex, world that is machine learning.

I am exceptionally grateful to all the individuals who gave up their time to participate as subjects in the project. In particular, I would like to acknowledge the courage and bravery of those who took part primarily in the hope that their personal experiences may one day help others. Furthermore, I would like to thank those staff of the South London and Maudsley National Health Service Trust for their invaluable help and collaboration with the recruitment process that made this study possible.

Lastly, I would like to thank my partner Akpene Akonou, for her unwavering love and kindness, my father, Gilbert Yeo, for his invaluable counsel and unshakable good spirit, and my mother, Helen Pettersson, whose years of support are beyond measure.

This thesis is dedicated to my mother, Helen Charlotte Pettersson

My Role in the Work Described

I am very grateful to my supervisors Dr Andrea Mechelli and Dr Paul Allen who devised the original study design. I am also grateful to my colleagues including, Stefania Benetti, with whom I collaborated in equal measure in order to i) recruit, assess and scan all participants included in the study, and ii) pre-process the neuroimaging and cognitive data in preparation for subsequent analysis; Dr Diana Prata, who provided the list of all single nucleotide polymorphisms reported by genome-wide association studies to date to be in association with psychosis; and Dr Andre Marquand, who provided me with both technical assistance and support for the multi-kernel learning analysis which could not be conducted using the PROBID graphical user interface. Alongside these procedures, I personally i) extracted the DNA from each subject's saliva sample allowing it to be sent away for genotyping, ii) performed all univariate and multivariate analyses described in the thesis, iii) chose, and implemented, the multivariate integration methods described in Chapter 5, and iv) interpreted and discussed all results detailed in the thesis. Clinical assessments were performed by a number of trained psychiatrists allied either to the Outreach and Support in South East London (OASIS) service with respect to At-Risk Mental State participants, or, with respect to first episode psychosis participants, any one of the early intervention in psychosis services offered by the South London and Maudsley (SLaM) National Health Services Trust located in Southwark, Lambeth, Lewisham or Croydon.

ABBREVIATIONS

ARMS	At-Risk Mental State
AV	Prediction Averaging
BSABS	Bonn Scale for the Assessment of Basic Symptoms
BSIP	Basel Screening Instrument for Psychosis
BSMCA	Best Single Modality Classification Accuracy
CAARMS	Comprehensive Assessment of At-Risk Mental States
ChSz	Chronic Schizophrenia
CVLT	California Verbal Learning Test
DTI	Diffusion Tensor Imaging
FA	Fractional Anisotropy
FEP	First Episode Psychosis
fMRI	Functional MRI
FRET	Fluorescent Resonance Energy Transfer system
GM	Grey Matter
HC	Healthy Control
HSCT	Hayling Sentence Completion Task
K	Kernel Matrix
K'	Integrated Kernel Matrix
KASP	KBiosciences Competitive Allele-Specific PCR
LOOCV	Leave-One-Out Cross Validation
MKL	Multi-Kernel Learning
MRI	Magnetic Resonance Imaging
MV	Majority Voting
OSH	Optimal Separating Hyperplane
PACE	Personal Assessment and Crisis Evaluation
PANSS	Positive and Negative Syndrome Scale
PCR	Polymerase Chain Reaction

SIPS	Structured Interview for Prodromal Symptoms
SK	'Simple' Sum of Kernels
sMRI	Structural MRI
SNP	Single Nucleotide Polymorphism
SVM	Support Vector Machine
T	Tesla
$\boldsymbol{w}$	Vector of predictive weights
WM	White Matter

Chapter 1

Background and Literature Review

1.1 Introduction

Schizophrenia (Sz) is a devastating illness that can have a profound impact on one's life (Gore et al., 2011). Both distressing and debilitating, it is characterised by a constellation of features that may be conceptualised in terms of four specific symptom dimensions; i) psychotic, positive, symptoms typically absent in the normal population (such as delusions, hallucinations and thought disorder), ii) alterations in volition and negative symptoms, typified by the absence of normal functioning (such as reduction in spontaneous speech and social withdrawal), iii) neurocognitive alterations (such as difficulty in memory, attention and executive functioning) and, iv) affective dysregulation (van Os & Kapur, 2009). A complex disorder, Sz has been defined for the purpose of clinical diagnosis, as the presence of 2 or more such symptoms (i.e. delusions, hallucinations, disorganised speech, grossly disorganised or catatonic behaviour, or negative symptoms) evident for a significant portion of time within a 1-month period, with continuous signs of disturbance for 6 months or longer, which result in a marked decline in social and/or occupational functioning, and which were not the direct result of a physiological substance or general medical condition, nor occurred concurrently with a depressive, manic, or mixed episode (American Psychiatric Association, 2000). With an estimated global annual incidence of 15.2 per 100,000 people, and lifetime morbid risk of 7.2 per 1000 people (McGrath et al., 2008), it is a

worldwide phenomenon that crosses both national borders and cultures (Jablensky, 2000).

Over the last century, a considerable body of work, encompassing a wide range of methodological approaches, has been generated with the aim of unravelling the potential neural, genetic and cognitive basis of the illness. Whilst this work has typically focused on patients with chronic schizophrenia (ChSz), more recently, many researchers have begun focusing their efforts on those thought to be in the illness' earliest stages, driven by the motivation to provide earlier and more effective treatment intervention. This relatively new line of research has focused primarily on those who have experienced a recent first episode of psychosis (FEP), and those assessed as having a significantly increased clinical risk of becoming psychotic, referred to as the 'At-Risk mental state' (ARMS). Coinciding with this shift furthermore, have been increasing calls for the results of psychosis-based research to be more readily translatable into the clinic (Borgwardt & Fusar-Poli, 2012; Matthews et al., 2006). As a result, one consequence has been the progressive development and usage of alternative analytical approaches that allow inference at the level of the individual, as opposed to more standard, (mass-) univariate, techniques which allow inference at the group level only. Based in this context, the overarching aims of the present doctoral work were as follows: i) to examine whether FEP and ARMS subjects could be differentiated at the group level from HCs, and each other, using standard univariate analyses of biological and cognitive data; ii) to investigate whether subjects with a FEP or ARMS may be identified at a the single-subject level using multivariate machine learning analysis of the same data; and iii) to identify whether the accuracy of the machine learning analysis could be enhanced by integrating different data types, and explore the optimal method

by which this might be achieved based on an empirical comparison of four different integrative approaches.

In what follows, I begin by providing an operational definition of FEP and the ARMS as used for clinical and research purposes, and within the current investigation, with the term psychosis referred to throughout in the context of schizophrenia specifically, unless otherwise stated. Next, I review the five types of data employed in this thesis, and the results obtained from previous studies when applied to the group level examination of ChSz patients relative to healthy controls (HCs), in addition to a summary of the findings when these techniques have been extended to FEP and ARMS subjects. I then give an outline exploring the multivariate machine learning analysis known as Support Vector Machine (SVM), which, with the ability to allow single-subject level inference, has emerged as a promising and increasingly used tool within the field of psychosis. Finally, I provide an explicit framework detailing the objectives for this work, in relation to the background context laid out below.

1.1.1 First Episode Psychosis

By definition, all individuals diagnosed with Sz will at some stage have had what is considered to be their first episode of psychosis, defined according to the severity, duration and context of their symptoms. Though investigations examining FEP subjects using techniques such as neuroimaging and advanced genetic analysis are relatively modern, the concept of investigating the defined FEP individual has been around for over thirty years (Gift et al., 1981). This was largely driven by the hope that examining FEP rather than ChSz patients, confounds such as exposure to anti-psychotic medication, effects of chronicity and effects of institutionalisation, would be reduced, if not wholly

absent, thus allowing the primary mechanisms underlying the illness to be more easily revealed. In accordance with the text-revised fourth edition of the Diagnostic and Statistics Manual of Mental Disorders (DSM-IV-TR), and as employed in the current investigation, a FEP meeting criteria for a schizophreniform psychosis requires that the patient presents with a combination of two or more positive (i.e. delusions, hallucinations, disorganised speech, grossly disorganised or catatonic behaviour) and/or negative symptoms (i.e. affective flattening, alogia and avolition) lasting between 1 and 6 months, which were not the direct physiological effect of a substance or general medical condition, and did not occur concurrently with a major depressive, manic, or mixed episode (American Psychiatric Association, 2000). In the event that symptoms persist for a period longer than 6 months, and occur alongside significant social and occupational decline, a confirmed diagnosis of Sz, as outlined above, is at this stage assigned. In such cases where a Sz diagnosis is made, studies suggest a substantial prospective consistency of up to 95% at one year (Addington et al., 2006). By investigating individuals who have only recently experienced a FEP therefore, it is ultimately hoped that treatment options can be developed that minimise and/or prevent the onset of established recurrent psychotic episodes in addition to the disabling decline in social and occupational functioning representative of ChSz (Keshavan & Schooler, 1992).

1.1.2 At-Risk Mental State

Building on previous work examining those with an increased familial/genetic risk of becoming psychotic, but who are, at present, clinically well (McGuffin et al., 1995), the ARMS is, in comparison, a relatively new category which places greater emphasis on measurable symptomatology over inherent genetic risk. Aimed at capturing young

people experiencing psychotic symptoms that are, clinically speaking, sub-threshold, this group are considered to have a significantly elevated risk ranging between 16-40% of developing a psychotic disorder within 1-3 years (Fusar-Poli et al., 2012a; Ruhrmann et al., 2005; Yung et al., 2008; Yung et al., 2003). Initiated by Yung and colleagues at the University of Melbourne, the primary objective of the ARMS classification was to identify those thought to be in the prodromal stage of psychosis with the aim of providing earlier and more effective treatment intervention thus preventing, delaying, and/or minimising, the likely first onset of frank psychosis (Yung et al., 1998). To identify those with an ARMS, an assessment tool was developed at the personal assessment and crisis evaluation (PACE) clinic (Yung, et al., 1998) known as the comprehensive assessment of at-risk mental states (CAARMS) (Phillips et al., 2000). Based on this assessment one is assigned as having an ARMS by meeting one of the following three criteria: 1) Attenuated psychotic symptoms not of psychotic intensity (e.g. ideas of reference, odd beliefs or magical thinking, perceptual disturbance and/or paranoid ideation) occurring several times a week for between one week and 5 years, 2) Brief Limited Intermittent Psychotic symptoms (e.g. the same symptoms described for category 1, but of psychotic intensity, which have a duration of less than one week, and resolve spontaneously within that time) and/or, 3) Trait and state risk factors combined with a significant decline in cognitive and social functioning over the past year (e.g. the individual has schizotypal personality disorder, or a first degree relative with a psychiatric disorder or schizotypal personality disorder, combined with a significant decrease in mental state functioning that occurred within the past year, lasting between one month and 5 years reflected by a 30 point reduction in the global assessment of functioning scale from premorbid level).

Since the inception of the CAARMS, a selection of other clinically based assessment tools devised for the same purpose have also been developed including the American Structured Interview for Prodromal Symptoms (SIPS) (Miller et al., 2002), the German Bonn Scale for the Assessment of Basic Symptoms (BSABS) (Klosterkötter et al., 2001) and the Swiss Basel Screening Instrument for Psychosis (BSIP) (Riecher-Rössler et al., 2008). The primary difference between the alternative measures tends to be the specific symptom type focused upon for assessment, with the CAARMS, SIPS and BSIP giving greater emphasis to attenuated positive symptoms, and the BSABS focusing more upon subjective cognitive disturbances, also referred to as basic symptoms (Schultze-Lutter, 2009). In spite of this, preliminary evidence suggests a good level of concordance between the different indices (Chuma & Mahadun, 2011). Since the ARMS subjects who participated in the current investigation were clinically assessed and identified using the CAARMS, however, it is this tool that I shall primarily focus on hereafter, though on occasions where previous studies have used genetic, or alternative clinical high-risk groups, this will be specified.

1.2 Previous studies of Chronic Schizophrenia, First Episode Psychosis and the At-Risk Mental State

As highlighted above, considerable effort has been made over the last century to understand the biological and cognitive pathological mechanisms underlying ChSz, which may, in turn, inform the identification of markers of psychosis risk. In more recent times, this effort has been further extended to studies involving FEP and ARMS subjects also. Within this framework, paralleling substantial developments in the fields of genetic profiling and magnetic resonance imaging (MRI) over the last three decades,

data obtained from genetic, neuroimaging and neuropsychological techniques have emerged as promising sources from which such markers may be drawn. This is reflected by results published to date reporting significant differences at group level, with respect to genotype, neuroanatomy, neurofunction and neuropsychological performance in ChSz, FEP, and ARMS subjects, relative to HCs, of which a summary review is provided below.

1.2.1 Genetics

1.2.1.1 Established Psychosis and Psychosis-Risk

The genetic contribution to Sz has been a topic of investigation since the beginning of the twentieth century, at which time it was widely believed to be a primarily hereditary disorder (McGuffin, et al., 1995). Such views were later challenged however, when studies demonstrated, for example, that genetically identical monozygotic twins could be discordant with respect to the illness (Stabenau & Pollin, 1967). After decades of similar efforts revolving around those with increased familial risk, in 2005, this investigative avenue was greatly advanced with the advent of the genome-wide association (GWA) study (Klein et al., 2005). Specifically, this largely automated technique allows an individual's entire genome to be searched and sequenced with respect to an expansive array of specific single nucleotide polymorphisms (SNPs) that commonly vary within the population. GWA studies therefore allow for an explorative approach with limited restriction placed on *a priori* hypothesised SNP involvement, making it well suited to the investigation of disorders thought to have a heterogeneous genetic basis. This is in contrast to disorders likely to be caused by a single gene for which alternative forms of analysis such as genetic linkage studies would be more appropriate. By comparing the frequencies with which particular SNPs occur between

different phenotypic subject populations (i.e. HCs and Sz patients), one can observe where the frequency of a given SNP is significantly higher in one population relative to another, in which case it is said to be 'associated' with that phenotype. To date, the results from GWA studies investigating ChSz suggest its genetic aetiology to be comprised of a large plethora of SNPs each with a relatively small effect individually, but magnified by the epistatic interactions between them (Burdick et al., 2008; Greenwood et al., 2011; Prata et al., 2009; Sei et al., 2010), which combine to confer an increased risk for the illness (Hansen et al., 2011; Ripke et al., 2011; Stefansson et al., 2009).

Similar to the sizable range of hypothesis-driven candidate genes reported by past studies to be in association with Sz (Egan et al., 2001; Hennah et al., 2003; Li et al., 2006; Williams et al., 2004), the corresponding genes for the plethora of SNPs implicated through more recent GWA studies are equally varied. These include for example, the zinc finger protein 804A (ZNF804A) gene (O'Donovan et al. 2008), the transcription factor 4 (TCF4) gene (Stefansson, et al. 2009), the Abelson helper integration-1 (AHI1) gene (Rivero et al., 2010) and the vaccinia-related kinase 2 (VRK2) gene (Steinberg et al., 2011) (for a complete list of all SNPs showing a positive genome-wide association with schizophrenia and/or bipolar disorder, published (or *in press*, to the best of my knowledge) until June 14th 2011 please see table 2.1). It is worth noting, however, that to date the majority of candidate genes previously reported to be in association with Sz have not been found to be significant by subsequent GWA analyses (Collins et al., 2012). Whilst a full discussion of the reasons underlying this discrepancy is beyond the remit of the current work, three potential explanations include: 1) that, in comparison to contemporary standards, the statistical power of previous hypothesis-driven candidate gene studies was relatively low thereby reducing

the probability of a gene implicated by a GWA study being similarly detected using a candidate gene approach, 2) the major hypotheses underpinning the selection of specific candidate genes were wrong, thus minimizing the probability that future GWA studies would replicate the candidate gene finding, and 3) the majority of hypothesis-driven candidate genes previously reported in the literature have only been assessed one to two times, resulting in an increased risk of false positives (for a more in-depth discussion investigating the discrepancies between the results of hypothesis-driven candidate gene studies and GWA studies please see Collins et al., 2012).

In addition to these findings a number of studies have also reported that, in some cases, known psychosis risk SNPs, and/or genotypes, may contribute to neuroanatomical and neurofunctional alterations often observed in established psychosis and which may also be evident prior to onset. Structural MRI (sMRI) studies for example have shown that in monozygotic twins discordant for Sz, total grey matter volume is significantly reduced in both twins relative to HCs (Borgwardt et al., 2010). Other studies have also suggested high genetic heritability rates with respect to brain volume demonstrating a significant association of reduced cerebrum and white matter volume with Sz risk genes (van Haren et al., 2012). Furthermore, in a recent meta-analysis, investigating specific neuroanatomical alterations in those at genetic high-risk of psychosis relative to HC and FEP subjects, reductions were reported in the right anterior cingulate and the left parahippocampal gyrus of genetic high-risk subjects versus HCs, and the right superior temporal gyrus, left insula, left cerebellum, and left anterior cingulate gyrus for FEP versus HCs. In addition, it was observed that when directly compared, FEP subjects showed significantly reduced grey matter in the precuneus, anterior cingulate, superior temporal gyrus and cerebellum relative to those at genetic high-risk. Based on these findings, the authors argued that reductions in the anterior cingulate are markers of

genetic liability to psychosis, whilst reductions in the superior temporal gyrus and cerebellum may represent markers of FEP onset (Fusar-Poli et al., 2012b). In accordance with these results, studies employing functional MRI (fMRI) have also reported significant main effects of psychosis associated genotype on brain activity during cognitive tasks. Pauli et al. for example published findings demonstrating an association between psychosis associated genes encoding for the dopamine transporter and the D-Amino acid oxidase activator and differential neural activation during a verbal fluency task (Pauli et al., 2012). Another fMRI study by Rasetti and colleagues further reported that the *ZNF804A* risk genotype for Sz was significantly associated with reduced functional connectivity between prefrontal and hippocampal regions in ChSz patients and their healthy siblings relative to healthy controls (HCs) (Rasetti et al., 2011). In a third study by Hall et al., the authors examined the impact of the NRG-1 risk allele SNP8NRG243177 in a cohort of subjects at genetically high-risk of Sz, due to having two or more close family members suffering from the disorder, but who were at the time of study, clinically well. Notably, the authors found the risk gene to be significantly associated with a range of measurable indices including, neural activation in fronto-temporal areas during a sentence completion task, the development and level of psychotic symptoms, and level of premorbid IQ (Hall et al., 2006). In addition to these findings, another more recent study has provided preliminary evidence that specific risk SNPs may even have the potential to be used as markers to help identify ARMS subjects, categorized using sub-clinical symptomatology rather than genetic risk, who will, and will not, transition to psychosis in the near future. Specifically, Kéri et al. genotyped 67 ARMS subjects with respect to the same NRG-1 SNP investigated by Hall et al., who were then followed up by clinicians, blinded to each subject's genetic data, for 12 months. Crucially, the authors found that of the 25 subjects who

carried the T/T risk allele genotype, all developed frank psychosis, compared to only 6, of the remaining 42 subjects, who did not have the T/T genotype. Taken together therefore, these studies in addition to others (Zinkstok et al., 2008), strongly support the proposition that a multitude of SNPs collectively result in significantly increased risk for the illness (Cichon et al., 2011; Ripke, et al., 2011; Stefansson, et al., 2009) and that in some cases, such genes may contribute to phenotypic alterations often seen in established psychosis, and which may also be evident prior to onset (Fusar-Poli, et al., 2012b; Hall, et al., 2006; Walton et al., 2012). This, in turn, supports the notion that specific risk SNPs/genotypes may potentially be used as markers to help identify those at psychosis-risk and those with established psychosis, with the tentative possibility arising from recent studies that specific SNPs may even be used to help predict which ARMS subjects, categorised using attenuated symptom criteria, are more, or less, likely to transition (Kéri et al., 2009).

1.2.2 Magnetic Resonance Imaging

Similarly to genetic research into Sz, substantial development in the field of MRI, and neuroimaging specifically, over the last 20-30 years has resulted in a considerable body of work that combined, supports the notion that alterations in brain wide neuroanatomy and function represent some of the most robust indicators of the illness available. Broadly speaking, neuroimaging studies may be divided into three categories, namely, those investigating grey matter (GM), those investigating white matter (WM), and those investigating neurofunction. In this context, whilst a range of different neuroimaging approaches are available for the investigation of Sz, including positron emission tomography (PET) and computerised axial tomography (CAT) for example, magnetic

resonance imaging (MRI) has emerged as one of the most commonly used techniques, largely driven by its non-invasive methodology and lack of radiation.

1.2.2.1 Structural MRI

1.2.2.1.1 Chronic Schizophrenia

To date a tremendous number of structural studies have been published reporting brain wide alterations, primarily reductions, in the GM of ChSz patients relative to HCs. One of the most common and well replicated of these findings for example is that of increased ventricular volume. Identified as early as the mid 1970's using CAT scanning (Johnstone et al., 1976), the finding has since been replicated by a host of subsequent sMRI investigations offering improved image resolution (Andreasen et al., 1990; James et al., 2002; Mathalon et al., 2001; van Haren et al., 2008). Since this promising start, a host of other sMRI studies have also been conducted identifying an array of *in vivo* GM changes in multiple cortical and subcortical regions. Whilst many of these employed a manual region of interest (ROI) approach, another analysis commonly used in recent times is voxel based morphometry (VBM). In brief VBM is an automated computerised technique that allows voxelwise analysis of structural brain images that, not requiring *a priori* defined regions of interest, can be conducted in a relatively exploratory fashion, providing estimates of group level difference(s) with an associated *p*-value (Ashburner & Friston, 2000). Using these univariate analyses, GM reductions in ChSz patients relative to HCs have to date been identified in multiple areas throughout the brain. Within the frontal lobe for example, Pomarol-Clotet and colleagues revealed GM reductions in the medial frontal cortex of 32 ChSz patients relative to matched HCs (Pomarol-Clotet et al., 2010). Similar findings were also reported by Kawada et al. and Kikinis et al. who reported GM reductions in ChSz patients relative to HCs in the

prefrontal cortex and middle frontal gyrus respectively (Kawada et al., 2009; Kikinis et al., 2010). Reductions have also been observed in the temporal lobe, with specific alterations reported in the posterior superior temporal gyrus (Menon et al., 1995), the inferior temporal gyrus (Ananth et al., 2002), and the middle and superior temporal gyri (Giuliani et al., 2005). Within the occipital lobe furthermore, reductions in GM have been noted within the cuneus (Neckelmann et al., 2006) and middle occipital gyrus (Ananth, et al., 2002; Cooke et al., 2008) of ChSz patients relative to HCs. Finally, GM reductions have also been observed in the parietal lobe, with a number of studies reporting reductions in the precuneus of patients relative to HCs (Antonova et al., 2005; Shapleske et al., 2002; Tanskanen et al., 2010). In addition to those cortical grey matter regions, alterations have also been identified in subcortical structures. Ananth et al. for example reported GM reductions in the mediodorsal thalamus (Ananth, et al., 2002), in line with similar findings by Tanskanen et al. who reported reductions not only in the thalamus but also posterior cingulate, parahippocampal gyri, and insula (Tanskanen, et al., 2010). Alongside these results, other studies have further extended the spatial findings, reporting various sub-cortical regions to be altered in ChSz patients relative to HCs, including the putamen, cerebellum, amygdala and hippocampus (Antonova, et al., 2005; Bonilha et al., 2008; Brown et al., 2011; Herold et al., 2009; Meisenzahl et al., 2008a). Taken together therefore, the data currently available suggests the presence of multiple regions of altered grey matter within both cortical and subcortical structures that could potentially be used as identifiable markers of patients with ChSz.

1.2.2.1.2 First Episode Psychosis and the At-Risk Mental State

Although fewer in number, sMRI studies of FEP and ARMS subjects conducted to date together suggest that the alterations in GM observed in the chronic stages of Sz appear

to be evident in early and prodromal psychosis as well, albeit to an often less severe degree (Egerton et al., 2011). Focusing on FEP subjects for example, a report by Chua and colleagues showed GM reduction in the prefrontal cortices in these patients relative to HCs (Chua et al., 2007). Similar frontal lobe findings have also been reported in the lateral prefrontal cortices (Kasparek et al., 2009), inferior frontal gyrus (Iwashiro et al., 2012), middle frontal gyrus (Job et al., 2002) and superior frontal gyrus in which notably, both decreases and increases have been observed (Molina et al., 2010; Price et al., 2010). Furthermore, and consistent with studies in ChSz patients, GM alterations in FEP subjects appear spatially extensive throughout the brain, with similar reports of, predominantly, reduced GM both in cortical regions, including superior (Kubicki et al., 2002; Price, et al., 2010), middle (Job, et al., 2002; Price, et al., 2010), and inferior (Molina, et al., 2010) areas of the temporal lobe, the cuneus area of the right occipital lobe (Molina, et al., 2010) and the left post-central and right inferior portion of the parietal lobe (Job, et al., 2002; Kubicki, et al., 2002), and subcortical regions, including the hippocampus, parahippocampus, anterior and posterior cingulate gyri, caudate nuclei and the thalamic nuclei (Chua, et al., 2007; Morgan et al., 2007; Salgado-Pineda et al., 2003).

Consistent with the findings from FEP studies, GM alterations have also been reported in multiple cortical and subcortical areas in subjects with an ARMS relative to HCs. Jung and colleagues for example recently reported GM reduction in the prefrontal, anterior cingulate, inferior parietal, parahippocampal and also superior temporal cortices of ARMS subjects relative to HCs (Jung et al., 2011). This observed plethora of localised GM alterations is also in line with observations made by Koutsouleris et al. who, based on ARMS subjects assessed using the BSABS and CAARMS combined, reported widespread GM reduction in an array of areas including the prefrontal cortices,

inferior and medial parietal cortices, posterior and medial temporal gyri, in addition to the insula, caudate nuclei, amygdala, hippocampal and parahippocampal cortices (Koutsouleris et al., 2009b). Similarly, Borgwardt and colleagues, whose ARMS subjects were categorised using the BSIP, reported widespread GM alteration including reductions in the posterior cingulate gyrus, precuneus, paracentral lobule, and superior parietal lobe, as well as an increase in the posterior temporal gyrus of subjects who later developed psychosis relative to HCs (Borgwardt et al., 2007a). Together the areas implicated by these three studies build on a small but growing number of reports suggesting widespread GM alteration in ARMS subjects relative to HCs, comprised predominantly of reductions but including instances of increase also (Borgwardt, et al., 2007a; Borgwardt et al., 2007b; Fusar-Poli et al., 2011b; Meisenzahl et al., 2008b; Stone et al., 2009).

In addition to investigations directly comparing FEP and ARMS subjects with HCs, a handful of studies have also tried to examine GM alterations specifically associated with conversion to psychosis. In a longitudinal study conducted by Pantelis and colleagues for example, they demonstrated that conversion to psychosis, relative to ARMS subjects who had not converted, was associated with GM reduction in the orbitofrontal cortex, cingulate gyrus, fusiform gyrus, left parahippocampal cortices and also the cerebellum (Pantelis et al., 2003). In another longitudinal study, whose ARMS subjects were in this case assessed using the BSIP, Borgwardt et al. reported that transition from the ARMS to FEP was associated with GM reductions in the orbitofrontal cortex, right inferior temporal gyrus, superior frontal gyrus, left precuneus and right cerebellar lobe (Borgwardt et al., 2008). In addition to this, two further studies by the same group, but this time using a direct cross-sectional comparison between different groups of ARMS and FEP subjects, found mixed results with one showing no

significant change between the two (Borgwardt, et al., 2007b), and the other reporting increased GM in the superior, middle and inferior temporal gyri bilaterally of ARMS relative to FEP subjects (Borgwardt, et al., 2007a). Collectively, these results are also consistent with another, more recent, multi-centre longitudinal study reported by Mechelli et al., in which the authors found significant GM reduction in the left parahippocampal cortex of ARMS subjects who later developed psychosis, relative to those who did not, whilst ARMS subjects generally had significantly less GM in frontal regions bilaterally relative to HCs (Mechelli, et al., 2011).

In summary, the results currently available strongly suggest FEP and the ARMS to be associated with significant GM alterations, predominantly reductions, observable at the group level. Notably, such changes do not appear to be restricted to any one region, occurring instead diffusely throughout both cortical and subcortical regions of the brain and mirroring those seen in ChSz albeit to a less severe degree. Furthermore, studies examining GM changes specifically associated with psychosis onset, suggest that FEP and the ARMS are each associated with a specific set of distinct, widespread differences, supporting the notion that, at least in terms of GM alteration, the ARMS and FEP cohorts may be treated as relatively discrete groups, with some alterations predating the expression of frank psychosis, and others only appearing post-onset.

1.2.2.2 Diffusion Tensor Imaging

1.2.2.2.1 Chronic Schizophrenia

Consistent with investigations of GM, many sMRI studies also report widespread alterations both in the volume and density of WM in the brains of ChSz patients relative to HCs, with a number of studies showing similar findings in FEP and ARMS subjects (Chua, et al., 2007; Cocchi et al., 2009; Di et al., 2009; Witthaus et al., 2008). However, progressive refinement in the sophistication of MRI technology over the last 20 years has recently enabled investigators to estimate the actual integrity of the WM network running throughout the brain as opposed to merely reporting increases or decreases in WM volume and/or density. Specifically, this has been made possible with the advent of diffusion tensor imaging (DTI). A type of diffusion weighted MRI, DTI uses the diffusion properties of water molecules within the brain in order to map out WM tracts, based on the premise that within these tracts the diffusion of molecules is restricted causing them to diffuse in an anisotropic way. This is in comparison to cerebro-spinal fluid (CSF) at the other diffusion extreme for example in which the diffusion of molecules is overwhelmingly isotropic. Using a diffusion tensor approach, whereby diffusion is represented in terms of an ellipse described by three orthogonal vectors, it is therefore possible to estimate the overall magnitude of the anisotropic diffusion within each voxel, and based on these, build a probabilistic network of all WM tracts in the brain (Jones, 2008). Whilst a number of diffusion indices are available using DTI, the one most commonly reported is fractional anisotropy (FA), which provides an estimate of the diffusivity within a voxel ranging from 0-1, and is considered to reflect the integrity of the underlying WM within that voxel. Hence, using this relatively novel method, and building on the results found using VBM and ROI analyses, a substantial cohort of DTI studies has now been published reporting WM integrity alterations in

ChSz patients relative to HCs. One recent study by Catani et al. for example identified significantly reduced FA values in the bilateral arcuate fasciculi of ChSz patients relative to HCs (Catani et al., 2011). Similarly, in another study by Knöchel and colleagues, the authors report reduced FA values in the inferior and superior genu and body of the corpus callosum, and also the isthmus (Knöchel et al., 2012) of ChSz patients relative to HCs. In a third recent publication by Kunimatsu et al., FA reductions were observed in a host of regions in patients relative to HCs, including both sides of the anterior cingulum, the cingulum body, uncinate fasciculus, the corpus callosum and also the left fornix (Kunimatsu et al., 2012). Together, these results are consistent with the data from forty previous DTI studies examining ChSz patients relative to HCs, summarised in a recent publication by Pettersson-Yeo et al. In accordance with previous findings, the review suggests ChSz to be primarily associated with reductions in WM integrity, evident in areas widely and diffusely spread throughout the brain including, the corpus callosum, anterior cingulate gyrus, arcuate, uncinate, fronto-occipital, superior and inferior longitudinal fasciculi, in addition to other frontal, temporal, parietal and occipital areas, and the cerebral peduncles (Ellison-Wright & Bullmore, 2009; Pettersson-Yeo et al., 2011).

1.2.2.2.2 First Episode Psychosis and the At-Risk Mental State

Though relatively fewer in number, studies have also been published examining DTI derived FA alterations in FEP and ARMS subjects. DTI studies of FEP for example have reported alterations, comprised predominantly of reductions, in a host of regions. For instance, in a recent paper by Carletti et al., WM reductions in FEP relative to HC subjects were reported in multiple areas including the left inferior longitudinal, fronto-occipital and superior longitudinal fasciculi, the posterior thalamic radiation, the

internal capsule, the posterior and superior corona radiata bilaterally, and the body and splenium of the corpus callosum (Carletti et al., 2012). These are mirrored by further studies which report FA reductions in the superior longitudinal and uncinate fasciculi bilaterally (Luck et al., 2010), splenium (Gasparotti et al., 2009), left fronto-occipital and inferior longitudinal fasciculi, right precuneus, right posterior limb of the internal capsule, and the left cerebral peduncle (Cheung et al., 2008), as well as one report of an increase found in the anterior cingulate gyrus (Segal et al., 2010), of FEP subjects relative to HCs. Inconsistent with these results however, a handful of studies also found no significant differences between FEP and HC subjects, including investigations by Friedman et al. (Friedman et al., 2008), Peters et al. (Peters et al., 2008) and two by Price and colleagues (Price et al., 2005; Price et al., 2008).

This mixture of results predominated by regional reductions, with a few instances of no difference at all, is mirrored by a small but growing number of DTI studies examining those deemed as having an ARMS. In this group for example, categorised in the following studies using the CAARMS or the SIPS, FA reductions have been reported in the left inferior and superior longitudinal and fronto-occipital fasciculi, the posterior thalamic radiation, the splenium and body of the corpus callosum (Carletti, et al., 2012), the superior longitudinal fasciculus (Karlsgodt et al., 2009) and superior frontal lobe (Bloemen et al., 2009) bilaterally, as well as in the right superior frontal lobe (Peters et al., 2009). Inconsistent with these results however, two investigations found no evidence of significant FA alteration in ARMS subjects relative to HCs (Peters, et al., 2008; Peters et al., 2010).

In addition to individual comparison with HCs, the study by Carletti et al., also reports a three-way cross-sectional comparison between FEP, ARMS and HC subjects. Notably,

the findings suggest that with respect to severity, the widespread WM alterations seen in the ARMS groups lie in an intermediate position between FEP and HCs (Carletti, et al., 2012), which, in accordance with GM results, supports the notion that the ARMS and FEP mental states are each associated with a set of relatively distinct WM alterations defined in terms of location and severity.

Based on the summary of studies presented here therefore, like GM, alterations in WM integrity, observed in multiple brain areas, appear evident as a likely neural correlate of those suffering ChSz. In comparison, though the picture with respect to FEP and ARMS subjects is less clear, the available data suggests a similar likelihood, with the majority of reports also showing FA reductions in multiple WM areas in these two groups. Importantly, as with GM findings, both the ARMS and FEP states appear to be associated with alterations in WM integrity that are relatively specific to each group. These inferences are in line with the review by Pettersson-Yeo et al., which also looked at DTI studies investigating FEP and ARMS subjects, from which three clear trends emerged. Firstly, where alterations in FA value, and therefore WM integrity, were observed these were predominantly decreases rather than increases. Secondly, alterations seemingly well established in ChSz appear largely mirrored, albeit to a less established degree in both FEP and ARMS subjects. Third and finally, where the alterations did occur, these were evident across a wide range of WM regions, with some areas reported particularly frequently including the frontal lobe, temporal lobe, corpus callosum, and anterior cingulate gyrus (Pettersson-Yeo, et al., 2011).

1.2.2.3 Functional MRI – Hayling Sentence Completion Task

1.2.2.3.1 Neuropsychological Performance

Consistent with the neuroanatomical data, a substantial body of literature is now available reporting significant group level differences in neurofunction in ChSz, FEP and ARMS subjects relative to HCs, detectable using functional MRI (Benetti et al., 2009; Broome et al., 2009; Crossley et al., 2009; Kindermann et al., 1997; Morey et al., 2005). Though such neural correlate differences have been observed across a range of cognitive dimensions, including working memory capacity, memory encoding and memory retrieval for example, for the purposes of this thesis, focus is given to the language initiation and inhibition components of executive function, as elicited by the Hayling sentence completion task (HSCT).

Originally devised as a neuropsychological test (Burgess, 1996), the HSCT requires subjects to complete a sentence in which the last word is missing, referred to as a sentence stem. Specifically, the task is divided into an initiation and inhibition condition with a balanced number of sentence stems in each. Whilst in the initiation condition, the subject must verbalise a word that is semantically congruent with the preceding sentence, in the inhibition condition in comparison, the word provided must be semantically and phonologically incongruent, forcing the subject to suppress the production of a congruent response (see section 2.4.2.2 for further detail).

Employing this task, a number of studies have reported that ChSz patients perform significantly worse when compared to HCs, particularly with respect to response inhibition. In 2006 for example, Chan and colleagues compared the scores of 90 ChSz subjects against the demographic norms provided with the test, reporting that for both initiation and inhibitory conditions, scores achieved by ChSz patients were significantly

worse than the associated norms by at least 1.5 standard deviations or more (Chan et al., 2006a). These findings were later built on by Joshua et al., who compared the performance of 39 ChSz patients with 44 matched HCs during the HSCT. In accordance with Chan and colleagues, they reported the performance of patients to be significantly worse than the control group for both initiation and suppression conditions (Joshua et al., 2009). Similar results have also been reported where the HSCT has been applied to FEP subjects. Using 78 FEP and 60 matched HC subjects for example, Chan et al. reported that FEP subjects made significantly more errors suggesting a deficit in their ability to initiate, or inhibit, a given response (Chan et al., 2006b). Taken together with the results of further studies examining ChSz patients (Chan et al., 2004; Groom et al., 2008; Marcezewski et al., 2001), the findings suggest that the HCST is a relatively robust measure of executive dysfunction in ChSz, and of FEP.

1.2.2.3.2 Neurofunction

At neurofunctional level, in conjunction with MRI, the HSCT has also been shown to robustly activate prefrontal and lateral temporal brain regions associated with language processing (Allen et al., 2008; Collette et al., 2001; Nathaniel-James, 1997). Based on this ability to activate executive and fronto-temporal networks, in 2004 Whalley et al. applied the task to a cohort of 69 subjects at genetic high-risk of psychosis, 27 of whom were experiencing isolated psychotic symptoms. In addition to 21 HCs, each subject underwent a functional MRI scan whilst performing a version of the HSCT in which they made covert (i.e. silent) responses. After the scan was over, subjects were provided with the same sentence stems as provided during the scan and asked to record the answers they had initially generated whilst being scanned. Interestingly, though the genetic high-risk group did not differ from HCs in terms of performance, the scan

suggested HCs had significantly greater activation during the task in a range of localised areas including medial prefrontal, thalamic and cerebellar regions (Whalley et al., 2004). In addition to this, later in 2005, the group re-analysed the same data using a functional connectivity analysis. Consistent with their original results, the new analysis revealed abnormally increased ipsalateral connectivity between left parietal and prefrontal cortical areas in the genetic high-risk group relative to HCs (Whalley et al., 2005). Building on these findings, in 2008 a complete version of the HSCT was adapted for fMRI by Allen and colleagues (Allen, et al., 2008). Significantly, the adapted task enabled subjects to overtly provide their answer to each sentence stem during the scan itself, allowing for a more direct measure of associated neural activity and behavioural performance. Subsequently, in 2010, the group applied this version of the HSCT to 15 ARMS and 15 HC subjects. Consistent with the findings in HCs reported by earlier studies, relative to baseline both groups showed significantly greater activation across a wide range of language associated regions, including the left superior, inferior and middle frontal gyri, the left middle temporal gyrus, the cuneus region of the occipital lobe and in the superior temporal pole bilaterally. However, the authors also reported a main effect of group showing ARMS subjects to have significantly increased activation relative to HCs in the right caudate and anterior cingulate gyri bilaterally, in addition to group by task results which showed that ARMS subjects had significantly increased activity in the anterior cingulate gyri bilaterally relative to HCs during the suppression condition specifically (Allen et al., 2010). In line with Whalley et al. however, interestingly, these alterations in neurofunction were observed in the absence of significant task performance difference between the two groups.

Using the HSCT adapted for fMRI, recent studies have also been conducted examining ChSz patients. Based on 19 ChSz patients and 12 matched HCs for example, Grosselin

et al. reported significant neural activation differences between the two groups characterised by areas of functional deactivation in the HC group which were absent in the patient group. In comparison to studies of ARMS and genetic-high risk subjects however, this difference was associated with significantly worse task performance in the ChSz group relative to the HC group (Grosselin et al., 2010). Consistent with these results furthermore are those of a later study, by Schneider et al., who used a region of interest approach focusing on areas of the default mode network. Specifically, they reported that during the HSCT, HCs exhibited significantly greater deactivation in the left lateral parietal region and posterior cingulate cortex relative to ChSz patients, and that this was specifically associated with slower reaction times and higher levels of performance error in the ChSz group (Schneider et al., 2011).

Taken together therefore, the results suggest that the HSCT is a relatively robust measure of executive impairment in ChSz and FEP patients. Furthermore, where this task has been used as a functional MRI paradigm, it has been shown to reliably activate executive and language networks within HC, established psychosis, and psychosis-risk groups respectively. Within these networks moreover, significant differences in neural activation and connectivity have been observed in ARMS, genetic high-risk, and ChSz subjects relative to HCs. Whilst these occurred in the absence of performance difference for ARMS and genetic high-risk subjects, interestingly, performance deficits were noted in the ChSz groups. From this combination of findings one possible inference is that although neurofunction of ARMS subjects is clearly altered relative to HCs, in comparison to FEP and ChSz subjects this alteration is not yet so severe as to impact their performance of the task. Collectively therefore, these results seem consistent with the reports of grey and white matter alteration already discussed, which suggest changes

similar to those seen in ChSz are also evident in ARMS and FEP subjects albeit to a less severe degree.

1.2.3 California Verbal Learning Test

1.2.3.1 Chronic Schizophrenia

In accordance with the neuroanatomical and neurofunctional data, patients with ChSz are associated with a wide range of neuropsychological deficits (Heinrichs & Zakzanis, 1998; Sponheim et al., 2010), many of which appear evident, albeit to less severe degree, in both FEP and ARMS subjects (Brewer et al., 2005; Mesholam-Gately et al., 2009; Pukrop et al., 2007; Wood et al., 2007a; Wood et al., 2007b). Of the different cognitive assessments currently used, one of the most frequently reported is the California Verbal Learning Test (CVLT). Designed to quantify different components of verbal learning, retention and retrieval (Delis et al., 1987), the CVLT is a neuropsychological test that comes provided with associated demographically corrected norms (Delis et al., 2000). Considering those studies that have used the CVLT with ChSz patients, overall findings suggest patients perform significantly worse than HCs across its range of dimensions. In a study involving 175 ChSz patients and 229 matched HCs for example, Paulsen et al. reported that patients performed significantly worse for all measures (Paulsen et al., 1995). Similarly, in a later investigation conducted by Tracy et al., it was reported that the performance of 28 ChSz patients was significantly worse than expected norms derived from the healthy population, with particular respect to learning and recall measures (Tracy et al., 2001). Together, these results are consistent with a seminal review conducted by Heinrichs and Zakanis examining neurocognitive deficits in ChSz. Specifically, though significant effect sizes reflecting broad cognitive impairment were observed for all 22 neurocognitive tests/constructs

assessed in the review, global and selective verbal memory of which the CVLT was used as a measure, had the first and tenth biggest effect sizes respectively (Heinrichs & Zakzanis, 1998).

1.2.3.2 First Episode Psychosis and the At-Risk Mental State

Deficits similar to those observed in ChSz have also been reported in instances where the CVLT has been applied to FEP and ARMS subjects. With respect to FEP subjects for example, Bilder et al. compared 94 FEP subjects with 36 matched controls, reporting that language and memory dysfunction, of which CVLT was used as a measure, best distinguished the two groups, with the patient group reported as having a generalised neuropsychological deficit of approximately 1.5 standard deviations relative to the HCs (Bilder et al., 2000). In a later study by Hill et al. comparing 62 FEP patients and 67 matched HCs, again, the patient group was reported to perform significantly worse on measures including verbal learning, short- and long-term memory and immediate recall, though no group differences were seen in the rate of forgetting or susceptibility to proactive, or retroactive interference (Hill et al., 2004). Consistent with these results, in another study by Friis et al. examining 219 FEP patients using the CVLT, amongst others, the authors reported the FEP group to have mean scores clearly below those associated with normal functioning (Friis et al., 2002). Together, these findings are in accordance with a recent meta-analysis examining neurocognition in FEP (Mesholam-Gately, et al., 2009). Including 43 separate samples comprising 2204 FEP subjects and 2775 age and gender matched HCs, the analysis demonstrated that with specific respect to the CVLT, FEP subjects differed from HCs on all aspects of the task. As seen in Heinrichs and Zakzanis, immediate verbal memory, a domain for which the CVLT was used as measure, again showed the largest effect size between FEP and

HC subjects relative to other cognitive domains examined. Furthermore, in accordance with data from other modalities, the authors report that the neuropsychological deficits observed in FEP, though not quite as severe, approach and broadly match those seen in ChSz (Mesholam-Gately, et al., 2009).

In accordance with findings from FEP subjects, similar deficits have also been observed at group level in those with an ARMS. Based on a cohort of 38 ARMS subjects, assessed using the SIPS, and 39 matched control subjects for example, using an assessment battery including the CVLT Lencz et al. showed verbal memory to be the most significantly impaired neurocognitive domain in the ARMS group relative to HCs (Lencz et al., 2006). Similar results were also reported by Hawkins et al. who performed the CVLT, amongst other tests, with 36 subjects also assessed as clinically high-risk based on the SIPS protocol. Interestingly, consistent with the trends observed for other modal approaches, the findings suggested that in terms of deficit severity, ARMS subjects lay in an intermediate position between HCs and more severely effected FEP subjects (Hawkins et al., 2004). This finding is also in accordance with results from a later, longitudinal, study conducted by Niendam et al. Whilst the team conducted an initial analysis examining 35 ARMS subjects, assessed using the SIPS, against matched normative data, they also performed a subsequent analysis comparing those subjects who after approximately 8 months, had further cognitive decline with those who had not. Consistent with previous studies, at baseline the team reported significant impairments in the ARMS group relative to HCs, with the results of the 8-month follow up suggesting a positive correlation between neurocognitive and functional stability (Niendam et al., 2007). Notably, this fits with the notion, observed from other modalities, that with respect to cognitive function there is a longitudinal decline in line

with progression from ARMS to FEP, which continues with subsequent decline to ChSz (Brewer, et al., 2005; Mesholam-Gately, et al., 2009).

In summary, consistent with the results of sMRI, DTI, and fMRI investigations, studies employing the CVLT suggest that patients with ChSz perform significantly worse relative to HCs across all dimensions of the task, with similar group level measures of deficit being mirrored in subjects with FEP and an ARMS, albeit on a declining scale of severity. This in turn supports the notion that CVLT outcome measures may potentially be used to help inform identification of those with established psychosis, in addition to those with an ARMS.

1.3 Support Vector Machine

As initially noted, the majority of work examining ChSz, FEP and ARMS subjects conducted to date, has primarily relied on traditional (mass-) univariate statistical analyses. Fundamentally, such analyses are geared toward finding gross focal abnormalities, either increases or decreases, which differ significantly at the group level between groups of interest. Though ideally suited for this purpose, such analyses do not allow inference to be made at the level of the individual, nor, are they able to take into account the inter-relationship between dependent variables. As suggested by the findings presented however, the alterations seen in those suffering from ChSz, FEP or the ARMS, tend to be multi-faceted, and, with respect to the neural correlates of these mental states furthermore, spatially diffuse. It follows therefore, that by employing a univariate analysis, potentially useful and insightful information encoded in the relationship between dependent variables is lost. Secondly, in spite of the numerous alterations revealed by standard univariate analyses across a range of biological, genetic

and cognitive dimensions with respect to ChSz, FEP and ARMS subjects, due to their group level nature, their potential for translation into a clinical setting has been limited, as it is a context in which decisions must be made about individuals.

Developed from the field of artificial intelligence, support vector machine (SVM) has been proposed as one possible analytical alternative that may circumvent the two limitations highlighted above (Brammer, 2009). A type of supervised machine learning pattern recognition algorithm, SVM is a multivariate approach able to classify subjects into predefined groups at the single subject level, subsequently providing a potentially high level of clinical translation. Its multivariate nature furthermore enables it to be sensitive to subtle, spatially distributed alterations that may otherwise be undetectable by an analogous univariate approach applied at the group level (Lao et al., 2004). Described more fully in section 2.5, in brief, the primary objective of SVM is to generate a decision function that can accurately classify single subjects into predefined groups, based on the totality of selected input data for each subject. It achieves this by representing the input data in kernel form transforming it from input space into a hyper dimensional feature space, in which an optimal separating hyperplane – representing the decision function – that is able to linearly separate the groups is found, which in input space may not be possible. By measuring the number of true, and false, positives, and negatives generated by the eventual classifier, an accuracy is obtained providing an estimate of how well it is likely to classify a new unseen subject (Schölkopf & Smola, 2002). Finally, an estimate of the classifier's statistical significance can be computed using a repeated cycle permutation test.

1.3.1 SVM applied to Neurological and Psychiatric disorders

Allowing inference to be made at the single subject level and providing a potentially high level of clinical translation, the use of SVM has become progressively widespread in the fields of both neurology and psychiatry, used in conjunction with a variety of different types of input data to accurately classify patients from HCs (Orrù et al., 2012). Using SVM in conjunction with sMRI for example a number of studies have reported being able to differentiate patients with probable dementia of Alzheimer's type (PDAT) from HCs with significant accuracies ranging from 82.7% to 94.5% (Arimura et al., 2008; Duchesne et al., 2008; Magnin et al., 2009; Nho et al., 2010; Oliveira et al., 2010). Providing results with even more direct clinical promise, several studies have also reported the ability to discriminate subjects with mild cognitive impairment, widely held to be the effective prodrome of established Alzheimer's disease, both from HCs and PDAT patients. Once again, these findings are notable for providing impressive classification accuracies, with values of up to 97.62% (Davatzikos et al., 2008; Gerardin et al., 2009; Plant et al., 2010). Still using sMRI in conjunction with SVM, similar results have also been obtained for a number of other mental disorders. For instance, Gong et al. reported being able to classify refractory and non-refractory depressive subjects from HCs using GM with accuracies of 67.39% and 76.09%, and WM with accuracies of 58.7% and 84.65% respectively (Gong et al., 2011). Similarly, Costafreda and colleagues reported successfully classifying patients with major depression from HCs with an accuracy of 67.6% (Costafreda et al., 2009a) based on whole brain neuroanatomy. Consistent with these results, patients with major depression have also been successfully discriminated from HCs based on fMRI data, with reported accuracies of 67.5% (Marquand et al., 2008) and 86% (Fu et al., 2008). DTI data furthermore has

also been used in conjunction with SVM to successfully discriminate patients from HCs, with one example of PDAT patients being successfully classified from HCs with 100% accuracy (Graña et al., 2011).

1.3.1.1 SVM applied to Chronic Schizophrenia, First Episode Psychosis and the At-Risk Mental State

Though relatively fewer in number, the results of the handful of SVM studies using single modality data to investigate ChSz, FEP and ARMS subjects have to date provided encouraging results. Using sMRI in conjunction with SVM for example, Davatzikos et al. reported differentiating ChSz patients from HCs with up to 81.1% classification accuracy (Davatzikos et al., 2005). Focusing instead on FEP subjects, Sun et al. also reported significant classification results, successfully discriminating FEP subjects from HCs with 86.1% accuracy (Sun et al., 2009). Consistent with these reports, sMRI in conjunction with SVM has also yielded similar results with respect to the ARMS group. For example, Koutsouleris et al. were able to successfully discriminate ARMS from HC subjects with 82% accuracy (Koutsouleris et al., 2009a). Alternatively, studies using DTI data in conjunction with SVM have also been published, with Ingalhalikar et al. for example reportedly able to differentiate ChSz patients from HCs based on patterns of WM alteration with up to 90.62% accuracy (Ingalhalikar et al., 2010). Using fMRI data furthermore, studies have also been performed successfully differentiating both ChSz and ARMS subjects from HCs. Based on patterns of neurofunction elicited during an auditory oddball task for example, Yang et al. reported being able to distinguish ChSz patients from HCs with 81.63% accuracy (Yang et al., 2010). Similarly, using patterns of neurofunction elicited during a verbal fluency task, Costafreda et al. were able to successfully distinguish ChSz patients from HCs with

92.4% accuracy (Costafreda et al., 2011). In addition to these studies, which used neuroimaging data in conjunction with SVM, results have also been published reporting successful classification accuracies using non-neuroimaging data. In 2010 for example, Yang et al. used a SVM approach in conjunction with genotypic data to successfully classify ChSz patients from HCs with 73.9% accuracy (Yang, et al., 2010), whilst in 2011, Koutsouleris and colleagues were reportedly able to discriminate ARMS subjects from HCs with up to 94.2% classification accuracy based on patterns of difference in their neuropsychological profile (Koutsouleris et al., 2011b).

Collectively therefore, the studies in which SVM has been applied to ChSz, FEP and ARMS subjects strongly suggest that this particular type of multivariate machine learning approach holds great potential with respect to helping identify members of each group at the single subject level. Moreover, the studies that are currently available encompass a wide range of data types, both neuroimaging and non-neuroimaging, for each of which SVM analysis was able to successfully identify patterns of alteration distinct to each group. Such findings are further reinforced by the results of comparable neurological and psychiatric studies where SVM has also been used, such as Alzheimer's disease, of which there are a greater number available (Orrù, et al., 2012). As such, the data presented here supports the notion that ChSz, FEP and the ARMS may be conceptualised as discrete mental states based on distinct patterns of biological and cognitive alteration, which may in turn be used to help inform identification of subjects from each group at the single-subject level.

1.4 SVM Integration

As reviewed here, the results currently available provide a strong case for the use of SVM to help inform identification of ChSz, FEP and ARMS subjects at individual level. In the context of developing SVM as a real world diagnostic aid however, it is arguable that greater, more consistent, levels of accuracy are required than those presently attained. One method proposed to achieve this is data integration. Based on the premise that SVMs trained using different data types will base their classifications on different patterns of alteration, as well as making different misclassifications, by integrating different data types it is intended that each data source will complement that of the other in order to enhance overall classification accuracy through the derivation of a consensus decision (Kittler et al., 1998). Broadly speaking, such derivation can be achieved in one of two ways, either 1) the kernel matrix for each individual data type can be integrated into a single new kernel matrix which simultaneously represents all data sources combined, which in turn can then be used to train a single integrated SVM, or 2) for each data type a single SVM classifier is trained, after which an ensemble decision is computed generating a single classification decision which is based on multiple SVMs, trained using multiple data types, combined.

In the context of mental illness, preliminary evidence from the study of Alzheimer's disease specifically, strongly supports such an approach with a small but growing number of studies reporting classification improvements based on the integration of different data types. Using a method known as multi kernel learning for example, which trains a single SVM using a kernel representing multiple data types, Hinrichs et al. reportedly increased their ability to differentiate PDAT from HC subjects to 92.4%, (Hinrichs et al., 2011), whilst Zhang et al. furthermore, using an approach similar to

multi kernel learning, reported a comparable integrated classification accuracy of 93.2% (Zhang et al., 2011). In a third study differentiating subjects with the postulated prodrome of PDAT, mild cognitive impairment, from HC subjects, Fan et al. were reportedly able to distinguish between subjects from the two groups with an unbiased integrated classification accuracy of 90%, again, using a single kernel, multiple data, SVM (Fan et al., 2008). Notably, these accuracies represented approximate increases of 7%, 5% and 3% respectively relative to the best single modality classification accuracy (BSMCA) in each case.

Such promising findings have recently now been extended to Sz research also. Specifically, in one recent paper by Yang et al., it was reported that by integrating both genetic and fMRI data, it was possible to distinguish ChSz patients from HCs with an accuracy of 87.25%, representing an approximate increase of 5% relative to the BSMCA (Yang, et al., 2010). In comparison to the those integrative studies aforementioned however, here the authors used a majority voting approach involving the generation of a single output classification decision based on an ensemble of multiple SVMs each trained using an individual data type.

Taken together, the preliminary evidence available supports the notion that integration per se has the potential to enhance classification accuracy beyond what is possible using single modality data alone. However, reflected by the differing methods employed by each of the studies cited, it remains unclear as to which type of integrative approach may provide the greatest classification enhancement and how this may be influenced by the number of data types integrated and/or the specific clinical comparison to which it is applied (e.g. FEP versus HC). Though not directly translatable to the clinical context here, clues to this question may initially be sought from the field of protein interactions.

Specifically, in 2006 Lewis et al. used heterogeneous data sets, each describing different protein properties, in conjunction with SVM, and compared the ability of two types of integrative technique to enhance classification accuracy relative to the BSMCA. The types of integration used included, i) an approach known as an unweighted 'simple' sum of kernels where the combined classification decision is based on a single SVM classifier trained using a single integrated kernel representing all data types combined, and ii) a version of multi kernel learning, which adds different weights to the data being integrated depending on its relative contribution to generating a successful classifier. Crucially, the authors suggest that whilst a relatively simple, un-weighted, averaging approach may be ideal when only two data types are being integrated, this may change to the more sophisticated multi kernel learning technique when three or more data sets are combined. Furthermore, they highlight the important fact that although integration overall may seem the optimum approach, in some instances the use of single modality data may in fact provide the highest classification accuracy achievable (Lewis et al., 2006).

1.5 Aims and Hypotheses

In summary, following the identification of gross group level differences between ChSz subjects and HCs using sMRI, DTI, fMRI, genotype and neuropsychological profile, a small but growing number of studies using single modality input data in conjunction with SVM, have shown that it is possible to classify patients from HCs with a significant degree of accuracy. Notably, this has been achieved using a range of different single modality data sources, both neuroimaging and non-neuroimaging. However, it remains less clear whether using the same metrics, it is possible to achieve

similar levels of accuracy in those with a FEP or ARMS. Furthermore, to the best of the author's knowledge, no study has yet conducted a cross sectional classification between FEP and ARMS subjects, and therefore it is unclear whether individuals from these two groups can be differentiated directly from each other at the single-subject level. Similarly, since no investigation has yet collected data from all five types of modality within the same study, the relative capacity of each data type to distinguish FEP and ARMS subjects from HCs, and each other, is also unknown.

In the context outlined above therefore, the primary objectives of the current thesis and the proposed hypotheses were as follows:

1. Employing a multimodal approach comprising sMRI, DTI, fMRI and neuropsychological data, in conjunction with standard univariate analyses, I aimed to examine whether group level differences were evident between FEP, ARMS and HC subjects in terms of neuroanatomy, neurofunction and neuropsychological performance. Based on previous results in both FEP and ARMS subjects relative to HCs, and each other, I hypothesised that at the group level, significant differences would be observable between:

i) FEP and HC subjects with respect to: a) GM, with reductions evident in multiple cortical and subcortical frontal, temporal and parietal regions of FEP subjects relative to HCs (see section 1.2.2.1.2), b) WM, with reductions evident in multiple areas including the inferior, superior and fronto-occipital fasciculi, the posterior thalamic radiations, the coronae radiatae, the internal capsule, and the corpus callosum of FEP subjects relative to HCs (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in parietal and posterior cingulate regions of HCs relative to FEP

subjects based on similar studies of ChSz patients (Schneider, et al., 2011), and, d) neuropsychological performance, with FEP subjects performing significantly worse for all subcomponents of the CVLT-II relative to HCs (see section 1.2.3.2).

ii) ARMS and HC subjects with respect to: a) GM, with reductions evident in cortical and subcortical frontal, temporal and parietal regions of the ARMS subjects relative to HCs (see section 1.2.2.1.2), b) WM, with reductions evident in the inferior, superior and fronto-occipital fasciculi, the posterior thalamic radiations, the coronae radiatae, the internal capsule and the corpus callosum of ARMS subjects relative to HCs (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in the right caudate and anterior cingulate gyri bilaterally of ARMS subjects relative to HCs (see section 1.2.2.3.2), and, d) neuropsychological performance, with ARMS subjects performing significantly worse for all subcomponents of the CVLT-II relative to HCs (see section 1.2.3.2).

iii) FEP and ARMS subjects with respect to: a) GM, with reductions evident in cortical and subcortical frontal, temporal, parietal and cerebellar regions of FEP relative to ARMS subjects (see section 1.2.2.1.2), b) WM, with reductions evident in the inferior, superior and fronto-occipital fasciculi, external and internal capsules, posterior thalamic radiations, coronae radiatae and corpus callosum of FEP relative to ARMS subjects (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in the anterior cingulate and caudate, in addition to increases in regions of the default mode network, in FEP relative to ARMS subjects, based on

previous results of ChSz patients relative to HCs, and ARMS subjects relative to HCs, respectively (see section 1.2.2.3.2), and, d) neuropsychological performance, with FEP subjects performing significantly worse than ARMS subjects for all subcomponents of the CVLT-II based on previous results showing progressive cognitive decline associated with conversion from ARMS to FEP (see section 1.2.3.2)

2. Using genetic, sMRI, DTI, fMRI and neuropsychological data, in conjunction with SVM, I aimed to identify which modalities could, and could not, accurately classify FEP and ARMS subjects at the single-subject level, with respect to both HCs and each other. Based on previous group level results suggesting that the biological and cognitive alterations observed in ARMS subjects relative to HCs are similar to, but less severe than, FEP subjects (Egerton, et al., 2011), I hypothesised that;

i) FEP subjects would be discriminable from HCs using the most types of data whilst in comparison, fewer data types would be able to discriminate ARMS from HC subjects. Specifically, I hypothesised that fMRI and genotype would be those data types least likely to successfully differentiate ARMS subjects from HCs, given that, a) with respect to neural activation, group level results for ARMS subjects have proved relatively less consistent with mixed reports of both increases and decreases relative to HCs, in addition to the magnitude of such differences typically less than the difference between FEP and HC subjects (Benetti, et al., 2009; Crossley, et al., 2009; Egerton, et al., 2011), and b) containing a proportion of individuals who will never transition to psychosis and who may also have no inherent familial risk, it is

logically less likely that a common pattern of SNPs associated specifically with established psychosis will be found that is shared by all ARMS subjects.

ii) FEP and ARMS subjects would be directly discriminable with structural MRI and neuropsychological data being the most able to differentiate the two, based on the relatively greater consistency and robustness of the group level findings of these two modalities when comparing FEP and ARMS subjects to HCs, in comparison to findings from fMRI, DTI and genetics studies (see section 1.2).

3. Thirdly, I aimed to elucidate whether classification accuracy may be enhanced through the integration of different types of data in order to generate a single overall classification output decision. To achieve this, I applied and compared four separate integrative methods to the same diagnostic comparisons, using combinations of two and three different data types. These integrative techniques comprised i) an un-weighted sum of kernels approach, ii) a multi-kernel learning approach, iii) an un-weighted averaging of prediction values, and iv) a majority voting approach. Based on similar studies in the field of Alzheimer's disease (Hinrichs, et al., 2011; Yang, et al., 2010; Zhang, et al., 2011), and also protein interaction prediction (Lewis, et al., 2006), I hypothesised that;

i) Classification accuracies would increase when integrating different types of data compared to when considering each type of data on its own

ii) For combinations of two data types, averaging and a sum of kernels approach would perform more successfully than a relatively more sophisticated multi kernel learning approach.

iii) For combinations of three data types, multi kernel learning would perform more successfully than either a sum of kernels, averaging or majority voting approach.

iv) The ability of each integration technique to enhance classification would vary with respect to the diagnostic comparison to which it was applied.

Chapter 2

Methodology

2.1 Participants

For a detailed characterisation of participant demographics, premorbid IQ and clinical data, please see Table 3.1 in section 3.1.2.

2.1.1 First Episode Psychosis

Nineteen first episode psychosis (FEP) subjects were recruited through the South London and Maudsley National Health Service Trust (www.slam.nhs.uk). All had experienced a first episode of psychosis within the past 24 months as determined by a trained psychiatrist that met DSM-IV-TR criteria for a schizophreniform psychosis. Specifically, each individual had presented with a combination of positive symptoms (e.g. hallucinations, delusions, disorganised speech) and/or negative symptoms (e.g. affective flattening, alogia, avolition) lasting between 1 and 6 months, which were not the direct physiological effects of a substance or general medical condition, and did not occur concurrently with a major depressive, manic or mixed episode (American Psychiatric Association, 2000).

2.1.2 At-Risk Mental State

Nineteen subjects with an At-Risk Mental State (ARMS) were recruited from OASIS (outreach and support in southeast London), a clinical service for young people aged 14-35 years old at high-risk of developing psychosis (Broome et al., 2005). Their status was assessed by a trained psychiatrist according to the PACE criteria (Yung, et al., 1998) using the CAARMS (comprehensive assessment of at-risk mental states) (Phillips, et al., 2000). In accordance with these criteria individuals were classed as having an ARMS based on the presence of 1) Attenuated psychotic symptoms not of psychotic intensity (e.g. ideas of reference, odd beliefs or magical thinking, perceptual disturbance and/or paranoid ideation) occurring several times a week for between one week and 5 years, 2) Brief Limited Intermittent Psychotic symptoms (e.g. the same symptoms described for category 1, but of psychotic intensity, which have a duration of less than one week, and resolve spontaneously within that time) and/or, 3) Trait and state risk factors combined with a significant decline in cognitive and social functioning over the past year (e.g. subject has a schizotypal personality disorder, or a first degree relative with a psychiatric disorder or schizotypal personality disorder, combined with a significant decrease in mental state functioning lasting between one month and 5 years reflected by a 30 point reduction in the global assessment of functioning scale from premorbid level).

2.1.3 Healthy Control

Twenty-three healthy control (HC) subjects were recruited from the local area through advertising. No HC subjects met criteria for a DSM-IV-TR psychiatric disorder, fulfilled the PACE criteria for prodromal symptoms nor had a first-degree family history of psychiatric disorder. Further, HCs were excluded if they had a current

medical illness, or had used any regular medication in the last 2 months.

2.1.4 Inclusion and Exclusion Criteria

Inclusion criteria required that all subjects speak English as their first language, have no history of neurological illness, drug, or, alcohol dependence, no significant visual or hearing impairment, and be aged 18-35 years old. Exclusion criteria for all subjects included a history of significant head trauma resulting in hospitalization and/or loss of consciousness, presence of a central nervous system disease, evidence of substance abuse or dependence disorder as defined by the DSM-IV-TR, and any contraindication to magnetic field exposure (e.g. metal implants, pregnancy or pacemaker).

2.1.5 Recruitment and Ethical Approval

The study was initially introduced to potential ARMS and FEP participants by their respective clinician. Subjects who were interested were then provided with an information sheet detailing the main aim of the study and an outline of the procedures involved. For all subjects, including HCs, a physical, or telephone, screening interview was arranged to discuss the study in greater detail and to confirm their eligibility. Before taking part, every subject was required to provide informed written consent. Throughout the recruitment process subjects were informed that they had the right to withdraw at anytime and without explanation. The study was conducted in accordance with ethical approval obtained from the National Health Service UK Research Ethics Committee (reference [08/H0805/64]) at the Institute of Psychiatry, King's College London, 16 De Crespigny Park, London, SE5 8AF.

2.1.6 Subject Pairing for Support Vector Machine

In order to use support vector machine (SVM) to investigate whether the three respective groups could be discriminated from each other, three binary comparisons for the three groups had to be created. These were ARMS versus HC subjects, FEP versus HC subjects, and FEP versus ARMS subjects. For each comparison, each subject from one group was required to be paired with an analogous subject in the corresponding comparator group, matched for age and gender. Using these criteria, both the ARMS versus HC, and FEP versus HC, comparisons comprised 19 subject pairs, and the FEP versus ARMS comparison 15 subject pairs from the total pool of subjects available matched for age (±4 years) and gender (see Table 3.1 in section 3.2.1 for a detailed characterisation of subject pairs). The rationale for subject pairing is explained in greater detail in section 2.5.3.

In order to allow a more direct comparison with the results of the multivariate analyses detailed in chapter 4, subject pairings were also maintained for standard analysis of the data (see chapter 3). This also allowed paired, rather than independent, samples t-test analyses to be performed providing the added benefit of enhanced sensitivity afforded by a paired t-test due to an intrinsically smaller standard error of the mean (Peers, 1996; Zimmerman, 1997).

2.1.7 Sample Size

2.1.7.1 Univariate Analysis

The number of subjects per group used for standard analysis in the current thesis are supported by a recent analysis of effect size in classic inference which suggests the optimum range to be 16-32 (Friston, 2012). In brief, whilst larger

samples are more likely to yield significant effects, there is a greater likelihood these will be of low, or trivial, size. Conversely, the same result detected with smaller subject numbers will be quantitatively stronger. The optimized range of 16-32 therefore maximizes sensitivity to large effects, whilst minimizing the risk of detecting trivial effects. The numbers used in the current study are further supported by the results from previous studies which, using similar subject numbers or fewer, were able to detect significant effects in both FEP and/or ARMS subjects relative to HCs, with respect to both measures of neuroanatomy (Borgwardt, et al., 2007a; Koutsouleris, et al., 2009b; Witthaus et al., 2009) and also neurofunction (Allen, et al., 2010; Benetti, et al., 2009; Crossley, et al., 2009).

2.1.7.2 Multivariate Analysis

The number of subjects used for multivariate analysis in the current study is consistent with past and ongoing empirical investigations at the Institute of Psychiatry which have suggested that accurate class discrimination in a clinical context is possible with as few as five or six subjects (Giampietro et al., 2010). Moreover, these numbers are further supported by previous machine learning studies of psychiatric and neurological illness which, using subject numbers similar to those here, have reported classification accuracies ranging between 70% and 100% discriminating individuals at the single-subject level, based on differing data types (Orrù, et al., 2012).

2.2 Pre-Scan Assessment

On the day of scanning, each subject was asked to attend the Institute of Psychiatry, where the research was conducted, two hours prior to scanning at which time they were assessed by a trained researcher, in order to quantify demographic, cognitive and clinical measurements for the subject. Following these assessments, each subject then entered the MRI scanner for a period of 1.5 – 2 hours during which time they underwent a series of structural and functional neuroimaging scans.

2.2.1 Demographic Assessment

During the pre-scan assessment period each subject was administered a structured interview in order to confirm their demographic particulars including age, gender and laterality, and to ensure compliance with the aforementioned inclusion and exclusion criteria. Laterality was determined using the Laterality Preference Inventory handedness subscale (Coren et al., 1979).

2.2.2 Neuropsychological Assessment

During the pre-scan assessment period each subject was administered the reading subtest of the wide ranging achievement test revised (WRAT-R) (Jastak & Wilkinson, 1984) and the California verbal learning test - second edition (CVLT-II) (Delis, et al., 2000).

2.2.2.1 Wide Ranging Achievement Test - Revised

Designed to assess a subject's ability to recognise and name letters and words (Jastak & Wilkinson, 1984) the reading subtest of the WRAT-R has been shown to provide an accurate estimate of premorbid IQ in clinical populations suffering brain dysfunction

(Johnstone et al., 1996) and is provided with associated demographically corrected norms. For each subject, their score was recorded and these were subsequently entered into a paired t-test using version 19 of the Statistical Package for the Social Sciences (SPSS) (www-01.ibm.com/software/analytics/spss/) to identify whether or not premorbid IQ differed significantly between any of the SVM comparator groups; ARMS versus HC, FEP versus HC and FEP versus ARMS subjects (see Table 3.1 in section 3.2.1 for a detailed characterisation of subject pairs).

2.2.2.2 California Verbal Learning Test

The CVLT-II is the second edition of a language and memory task designed to quantify an individual's ability for verbal learning, retention and retrieval (Delis, et al., 1987) and is provided with associated demographically corrected norms (Delis, et al., 2000). During administration of the test each subject's responses are recorded manually, and subsequently inputted into the associated CVLT-II software package (Delis & Fridlund, 2000). This software provides a comprehensive summary of raw and demographically corrected standardised scores for the different task components. To allow analysis using SVM, in accordance with the steps outlined in section 2.5, CVLT data must be prepared for input into SVM, i.e. feature extraction. This is achieved by each score, comprising the totality of different components, being collated into a single column vector for each subject. Alternatively, in order to conduct a standard analysis, the scores for each component of the task can be entered into a standard paired t-test performed in SPSS19 to identify the presence of group differences with respect to any of the components.

2.2.3 Clinical Assessment

During the pre-scan assessment period each subject was administered the Positive and Negative Syndrome Scale (PANSS) for schizophrenia (Kay, 1987). The PANSS is used to quantify, in a given individual, the degree of positive symptoms present; referring to an excess or distortion of normal function (e.g. hallucinations and delusions), the degree of negative symptoms present; referring to the loss or reduction of normal function (e.g. blunted affect and emotional withdrawal) and/or the degree of general psychopathology present (e.g. unusual thought content or active social avoidance). The interview is divided into 7 positive-symptom items, 7 negative-symptom items and 16 general psychopathology symptom items, in order to provide a balanced representation of each symptom type, and also their inter-relationship, with each item scored on a seven-point severity scale ranging from absent (1) to extreme (7).

Widely used in the assessment of schizophrenia and psychosis, the PANSS has been shown to provide a good level of construct validity and inter-rater reliability (Kay, 1987).

2.3 Molecular Genetics

2.3.1 Collection

Saliva samples were obtained from each subject with informed consent, using the Oragene® DNA collection kit (DNA Genotek Inc, Ontario, Canada), preceded by half an hour of nil by mouth. DNA was manually extracted as per the recommended protocol established by DNA Genotek Incorporated (www.dnagenotek.com/ROW/pdf/PD-PR-006.pdf). Extracted DNA was then genotyped for a pre-selected list of psychosis

associated SNPs (see Table 2.1). This is a complete list of all SNPs showing a positive genome-wide association with schizophrenia and/or bipolar disorder, published (or *in press*, to the best of my knowledge) until June 14th 2011. Genotyping was performed by KBioscience (www.kbioscience.co.uk), using KASP SNP genotyping system, a homogenous fluorescent resonance energy transfer (FRET) based system, coupled with competitive allele specific PCR. For more detailed information, see the KASP SNP genotyping manual at www.kbioscience.co.uk/download/KASP%20Manual%20V4%200.pdf.

2.3.2 Preparation of Genetic data for input into SVM

In accordance with the steps outlined in section 2.5, it was necessary to prepare the genetic data for input into SVM by encoding the totality of data into a single column vector. To ensure that any SVM classifier generated using these data were not influenced by artificial weighting of the different genotypes due to ordinal or scalar coding, the genotype of each SNP for each subject had to be orthogonally coded – e.g. the homozygous/heterozygous genotypes AA, AB, BB for a given SNP could be coded '1 0 0', '0 1 0' and '0 0 1'. The SNP values for each subject were then collated into a single column vector that could be entered into a SVM. In cases where one or more SNPs could not be genotyped for a particular subject, these had to be excluded for all other pairs in the SVM comparison since the length of each column vector entered must be the same for each subject and group, which is also true for the data length of any other type entered into a SVM (for further information see 2.5.1).

Table 2.1. Table showing the 26 SNPs selected for SVM input and the corresponding publication from which they were derived

Gene	**Single Nucleotide Polymorphism**	
1. ZNF804A	rs1344706	(O'Donovan et al., 2008)
2. CACNA1C	rs1006737	(Ferreira et al., 2008)
3. MHC/PRSS	rs13211507	(Steinberg et al., 2011)
4. TCF4	rs9960767	(Stefansson, et al., 2009)
5. MMP16	rs7004633	(Ripke, et al., 2011)
6. NRGN	rs12807809	(Stefansson, et al., 2009)
7. CMYA5	rs10043986	(Chen et al., 2011)
8. CMYA5	rs4704591	(Chen, et al., 2011)
9. MHC/PRSS	rs3131296	(Stefansson, et al., 2009)
10. MHC/PRSS	rs6932590	(Stefansson, et al., 2009)
11. NCAN	rs1064395	(Cichon, et al., 2011)
12. PBRM1	rs2251219	(Williams et al., 2011)
13. TCF4/CCDC68	rs4309482	(Steinberg, et al., 2011)
14. AHIL	rs7750586	(Rivero et al., 2010)
15. MHC/PRSS	rs911507	(Steinberg, et al., 2011)
16. PCGEM1	rs17662626	(Steinberg, et al., 2011)
17. CNNM2	rs7914558	(Ripke, et al., 2011)
18. NT5C2	rs11191580	(Ripke, et al., 2011)
19. ANK3	rs10994336	(Ferreira, et al., 2008)
20. ANK3	rs9804190	(Schulze et al., 2009)
21. CSMD1	rs10503253	(Ripke, et al., 2011)
22. TCF7L2	rs7903146	(Hansen, et al., 2011)
23. VRK2	rs2312147	(Steinberg, et al., 2011)
24. CACNA1C	rs7972947	(Ripke, et al., 2011)
25. DYPD	rs1625579	(Ripke, et al., 2011)
26. TRIM26	rs2021722	(Ripke, et al., 2011)

ZNF804A: Zinc finger protein 804A, CACNA1C: Calcium channel, voltage dependent, L-type, alpha 1 subunit, MHC/PRSS: Major histocompatibility complex/Cationic trypsinogen gene, TCF4: Transcription factor 4, MMP16: Matrix metallopeptidase 16, NRGN: Neurogranin, CMYA5: Cardiomyopathy associated 5, NCAN: Neurocan, PBRM1: Protein polybromo1, CCDC68: Coiled coil domain containing 68, AHIL: Abelson helper integration 1, PCGEM1: Prostate specific transcript 1, CNNM2: Cyclin M2, NT5C2: 5'-Nucleotidase cytosolic II, ANK3: Ankyrin 3, CSMD1: CUB and Sushi multiple domains 1, TCF7L2: Transcription factor 7-like-2, VRK2: Vaccinia related kinase 2, DYPD: dihydropyrimidine dehyrdrogenase, TRIM26: Tripartite motif containing 26.

2.4 Magnetic Resonance Imaging

2.4.1 Structural Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a technique that can be used to visualise different tissues of the human body in a non-invasive manner, utilising properties of the protons contained within the nuclei of hydrogen atoms found in every cell of the body.

Within each hydrogen atom the charged proton rotates, or 'spins', on its own axis with a specific direction and intensity referred to as angular momentum. This in turn creates a small magnetic field known as the magnetic moment. Although every proton has angular momentum, in atoms that have equal numbers of protons and neutrons the magnetic momentum created by each cancels that of the other. In those atoms with unequal numbers in contrast, such as hydrogen which only has a single proton in its nucleus, there is a net magnetisation.

In the absence of an external magnetic field, the protons within each hydrogen nuclei are randomly aligned with the net magnetisation effectively equal to zero. Placing the hydrogen nuclei into an external magnetic field (B_0) however forces the protons into alignment in the direction of the field. The interaction between the external field and the magnetic moment of the proton also causes the proton to rotate about the longitudinal (z) axis of B_0 in a process referred to as precession (see Fig. 2.1a), at a frequency known as the Larmor frequency; governed in part by the strength of the external magnetic field.

Figure 2.1. a) b)

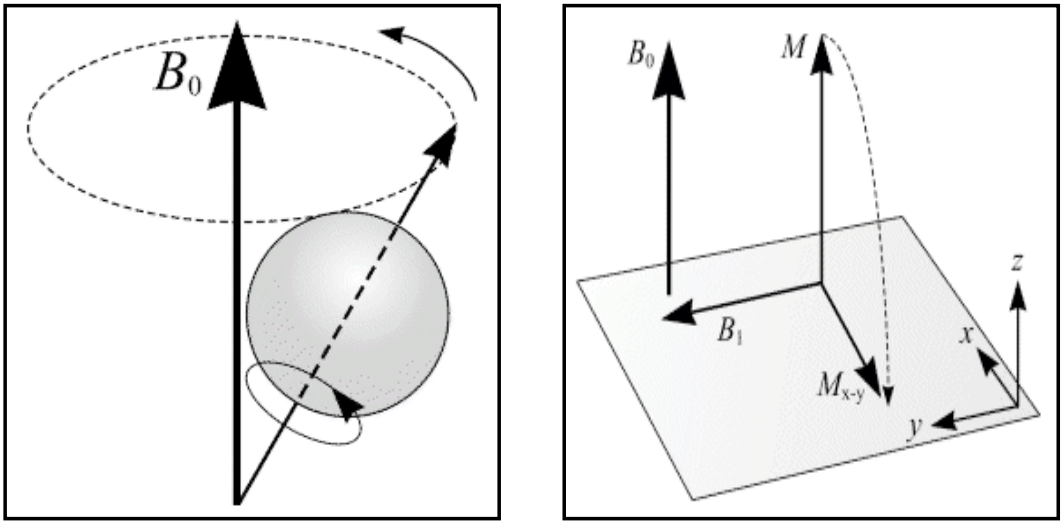


Figure 2.1 a-b adapted from (Puddephat, 2010). Fig. 2.1a represents a proton both spinning on its own axis and precessing about the z-axis of the external magnetic field gradient B_0. Fig. 2.1b demonstrates the principle of protons being 'flipped' into the transverse x-y plane by a resonant radio frequency pulse which generates a magnetic field B_1 perpendicular to B_0

Within the external magnetic field protons are aligned in either a low-, or high-energy state, which are said to be parallel and anti-parallel to the longitudinal z-axis of the field respectively. Here a Cartesian x, y, z coordinate system is used whereby the 'spins' of each proton are described with respect to a stationary frame of reference. Without further interference the overall number of protons in the low energy parallel state outnumber those in the high energy anti-parallel state resulting in a net magnetisation vector (NMV or M_0) in alignment with the longitudinal z-axis of the external magnetic field, referred to as longitudinal magnetisation.

By applying a radio frequency (RF) pulse that is at the same, or resonant, frequency at which the protons are precessing, it is possible to force protons in parallel alignment, into the excited, higher energy anti-parallel state, resulting in the NMV now being in the transverse plane. The pulse also creates a second external magnetic field (B_1), which,

applied perpendicular to the z-axis, causes the protons to precess with the same frequency within the transverse x-y plane, at which stage they are said to be in phase coherence (see figure 2.1b). This rotating magnetic field, precessing about the z-axis in the transverse plane, in turn generates a measurable alternate current equal to the RF pulse which can be detected using a receiver coil. The angle through which the NMV is rotated away from the z-axis is known as the flip angle and is proportional to the strength of the B_1 field generated by the RF pulse. Once the RF pulse is removed, the protons that have 'flipped' into the transverse plane begin to realign themselves with the longitudinal z-axis as they return to their lower original energy state (longitudinal relaxation) and simultaneously, phase coherence is progressively lost as the NMV in the transverse plane diminishes (transverse relaxation).

Relying on these principles, an MRI scanner functions by creating a strong static magnetic field B_0, measured in Tesla (T) units, at the centre of which is placed the area of interest to be scanned. Once the protons have been given time to align with the z-axis of B_0, an RF pulse is applied using a transmission coil which flips the NMV into the transverse plane. The alternate current, generated by the NMV now rotating about the B_0 axis in the transverse plane, is then recorded by a receiver coil and is known as the magnetic resonance (MR) signal. The time taken between the RF pulse being applied and an MR signal being received is known as the echo time (TE). Similarly, the time between the application of each RF pulse is known at the repetition time (TR).

The rate at which the NMV returns to the longitudinal from the transverse plane is exponential and can be defined by the time constant T1. This refers to the time taken for the longitudinal NMV to return to 63% of its original equilibrium value under B_0 following a 90° pulse, and occurs as the protons excited by the RF pulse omit thermal energy to the surrounding lattice resulting in their return to the longitudinal plane. On

images where a T1 weighting is used, those tissues with a long T1 such as fluid, have low signal intensity and therefore appear dark on the image, whilst those with a short T1 such as fat (and grey matter) have a high signal intensity and therefore appear bright on the image. Conversely the rate at which the protons in the transverse plane lose their phase coherence, resulting in the loss of the NMV in that plane, can be represented by an exponential decay function with a time constant T2. In contrast this refers to the time taken for the NMV in the transverse plane to decay to approximately 37% of its original value, as a result of the 'spins' interacting with each other. The rate of decay in phase coherence is also affected by inhomogeneity within the external magnetic field however, resulting in a faster loss of phase coherence, and since even in the most modern scanners inhomogeneity is found in the static magnetic field, this is taken into account by use of the time constant T2*. Opposite to T1, in T2 (or T2*) weighted images, tissues with a long T2 such as fluids appear brighter on the image, versus tissues with a shorter T2 such as fat, or grey matter, which appear darker.

2.4.1.1 Image formation

Relying on principles outlined above, MR images are generated in three stages through a series of different frequency RF pulses applied at different times, namely, 1) slice selection, 2) frequency encoding and 3) phase encoding.

2.4.1.2 Slice Selection

During the first stage, slice selection, an additional magnetic field gradient is superimposed around the B_0 field with the longitudinal axis orientated in the z-direction, with the strength of the field graded along the axis. This variation in field strength causes the protons to precess at differing frequencies proportional to the distance of the

nuclei from the centre of the superimposed magnetic field. When a single frequency RF pulse is then applied, only those protons located at the centre representing a narrow plane perpendicular to the longitudinal axis, precessing at the same frequency as the RF pulse, will absorb the RF energy and thus be excited into the transverse plane. This allows a given slice along the z-axis to be selected, and its thickness determined by the strength of the superimposed magnetic field gradient.

2.4.1.3 Frequency Encoding

During frequency encoding, a different magnetic field gradient is superimposed upon B_0 such that the precessional frequency of protons is graded along the x-axis. Application of an RF pulse and subsequent recording of the MR signal results in spatial encoding along the x-axis, reflecting the interference pattern formed by the different frequencies along the x-axis.

2.4.1.4 Phase Encoding

As with slice selection, phase encoding requires the application of a magnetic field gradient superimposed upon B_0, and is used to take account of the alterations in phases that differ along the gradient applied in frequency encoding. Here, the gradient is applied along the y-plane orthogonal to those used in slice selection and frequency encoding. A pulse sequence is then repeatedly applied with only the phase encoding gradient changing, with field strength declining to zero and then increasing back to its original amplitude. The number of times the pulse sequence is repeated is equal to the number of pixels in the subsequent image matrix generated.

Repeating phase and frequency encoding for each slice along the z-axis, corresponding information is collected for each volume-element, or voxel, the size of which is governed by the slice selection gradient. A Fourier transformation can then be used to generate a signal intensity for each, based on the phase and frequency information gathered during encoding, which in turn can be converted into intensities on a grey scale forming a 3D volumetric image, the resolution of which depends on the voxel size.

2.4.1.5 Structural MRI sequence

In the current study a 3D Spoiled Gradient Recalled Echo (SPGR) pulse sequence was used to create a structural T1 weighted image for each subject. This sequence comprises a very short TE, TR and a small flip angle. The short TE results in the T1 image being created in the absence of interference from transverse relaxation. To ensure this absence, a further magnetic field gradient is applied shortly after the slice selection gradient which 'spoils' the transverse magnetisation vector, by heightening the speed of the natural dephasing of the proton coherence in the x-y plane, thus eliminating it. Further, the sequence allows for very thin slices to be selected, resulting in an image of high resolution that due to voxels being isotropic can be viewed in the x, y or z planes.

2.4.1.6 Structural MRI Analysis

There are currently a wide range of toolboxes that implement the analysis of structural MRI data. For the purpose of the present project I used voxel based morphometry (VBM) as implemented in the SPM8 (www.fil.ion.ucl.ac.uk/spm) toolbox.

2.4.1.6.1 Voxel Based Morphometry

Voxel based morphometry (VBM) is a technique developed by Ashburner and Friston with the objective of identifying differences in neuroanatomy between groups (Ashburner & Friston, 2000). Prior to its development, identification of group level neuroanatomical difference relied on manual examination of a specific brain structure (regions of interest) across subjects. As an automated computational approach, VBM improved upon this method by removing inconsistencies in inter-rater reliability and consistency, and reducing the associated time and labour intensity. Simply the core function of VBM is to compare tissue densities/volumes on a voxel by voxel basis across the whole brain using deformation fields, thus omitting the need for an *a priori* hypothesised region of interest. To achieve this, the high-resolution structural scan for each subject first undergoes a series of three preprocessing steps; normalisation, segmentation and smoothing.

The normalisation step is employed in order that the brain image for each subject is warped to a template image, typically the Montreal Neurological Institute-305 (MNI-305) template, altering the overall size and shape of each brain so that it is comparable to the template, placing it in the same standard space as every other brain allowing for group level comparison. During the subsequent segmentation step, based on the signal intensity, and prior probability tissue maps, the image for each subject is segmented into compartments of grey matter, white matter and cerebrospinal fluid facilitating subsequent analysis focusing on one, or a combination, of these different tissue types. Recently, to refine the process these two steps have been combined into the unified segmentation procedure (Ashburner & Friston, 2005) which is used in combination with the Diffeomorphic Anatomical Registration through the Exponentiated Lie algebra (DARTEL) algorithm (Ashburner, 2007) to segment and subsequently normalise the

reoriented structural images. This is achieved first, by the creation of a study specific template derived from the mean of all subjects' images. The deformations from this initial custom template to the image of each individual subject are then computed and the inverse of these deformations applied to each individual image, which are then used to generate a new custom template. This process is iteratively repeated a set number of times, and the final deformation parameters used to warp the image of each subject with a high level of precision to the same standard space. In contrast to using a pre-generated template (e.g. MNI-305), this technique allows each subject's image to be segmented with greater accuracy and sensitivity (Yassa & Stark, 2009). An additional "modulation" step can also implemented to account for the expansions and/or contractions of brain regions during the normalisation process that in turn alter the volumetric measurements for each voxel. This consists of multiplying the spatially normalised grey matter after normalisation, by the volumetric changes encoded during the warping procedure implemented for normalisation, so as to ensure that the total amount of signal in each voxel is conserved.

The final step of smoothing involves convolving an isotropic kernel with the image in order to account for gross inter-subject variability, by averaging the signal intensity over a given region defined by the size of the kernel such that each voxel represents the average signal intensity for that region. The primary objectives of this step are three fold. First, it allows the confident application of parametric statistical tests by rendering the data distribution more parametric. Second, it is required to comply with the assumptions underlying Gaussian Field theory. Third and finally, the size of kernel used can be altered to allow averaging across subjects by accounting for regional alterations in structure that occur at a spatial scale which has corresponding homologous anatomical

regions that exist for all subjects. Together, these transform the images enabling them to be analysed using voxelwise statistics.

2.4.1.6.2 Standard Univariate Analysis

To perform a standard univariate analysis of the images, the general linear model (GLM) is employed (Frackowiak et al., 2004). This is a flexible framework which allows the implementation of a range of inter-group statistical tests to each voxel in turn, based on the underlying assumption that the data can be described in terms of effects of interest, confounds of no interest, and a residual error variable. In the statistical analysis, a design matrix is used to fit a pre-defined model to the observed data in order to estimate the contribution, referred to as parameter weights, of each variable of interest to the observed data. Standard parametric tests can then be applied to these parameter estimates (e.g. t-test, F-test, ANOVA) to identify differences between effects of interest at the voxel level. The results can then be used to generate a statistical parametric map (SPM) with each voxel given a t-statistic. Due to the mass-univariate nature of the approach, a correction for multiple comparisons based on Gaussian random field (GRF) theory is also applied which accounts for the number of voxels considered whilst minimising the risk of false positives through consideration of the likelihood of false positives in a random data set (Worsley et al., 1996). The fact that the data points are not independent of each other means that a standard Bonferroni correction cannot be used as it would be too conservative an approach.

2.4.2 Functional Magnetic Resonance Imaging

Functional MRI (fMRI) is a non-invasive imaging technique used to spatially map brain activity occurring during a given task. Developed in the 1990s (Ogawa et al., 1990a),

fMRI relies on the differing magnetic properties of oxygenated and de-oxygenated haemoglobin (HGb) found in red blood cells (RBCs). Specifically deoxyhaemoglobin is paramagnetic in nature - unlike oxyhaemoglobin which is diamagnetic - and in its presence it has been shown that the MR signal detected from protons in the water molecules of tissue surrounding blood vessels is altered (Ogawa & Lee, 1990; Ogawa et al., 1990b). This alteration creates a natural blood oxygen level dependent (BOLD) contrast in the MR signal, particularly T2, in these tissues relative to the surrounding tissues. Due to neurovascular coupling, in periods of increased neuronal activity, blood circulation increases to active areas consequent to an increased need for oxygen proportionate to the level of activity. The dynamic changes that occur during this coupling are referred to as the haemodynamic response. In combination with specifically designed magnetic field gradient and RF pulse sequences, the BOLD contrast can be used as a proximal measure of neural activity over time. Employing the principles of MR signal detection described above, a grey scale 'functional' image can thus be created for each time point of an MRI scan. Together, these scans can be used to image the haemodynamic response with respect to time in each area of the brain during a given task, allowing inference to be made regarding the areas involved during the associated mental processes.

The haemodynamic response and its relation to subsequent image generation can be described in three phases; 1) areas of neuronal activity will have an initial decline in image intensity associated with the consumption of oxygen (through neuronal respiration) resulting in increased levels of paramagnetic deoxyhaemoglobin (Vanzetta et al., 2004), 2) in response to this loss, vascular flow loaded with oxyhaemoglobin increases to these areas, resulting in an excess of diamagnetic haemoglobin which

generates an increase in signal intensity (Buxton et al., 1998), and 3) termination of the signal alteration as the excess oxyhaemoglobin is used or removed.

Work conducted by Logothetis et al. (Logothetis & Pfeuffer, 2004) has further demonstrated that the electrical signal generated by neuronal activity correlates well with dynamic changes in oxygen demand and supply. In combination with the neurovascular coupling described above representing the intrinsic link between areas of neuronal activity and increased blood flow (Ogawa et al., 1993), BOLD contrast can hence be used to accurately image localised areas of neuronal activity.

2.4.2.1 Functional MRI sequence

Functional MRI images were acquired for the present study using a technique known as echo-planar imaging (EPI) (Mansfield, 1984). This approach uses the switching between very strong magnetic field gradients to collect T2* weighted data for the entire 2D slice in one go, prior to the T2* decay, allowing rapid acquisition of the entire image volume (usually covering the whole brain). This high speed of acquisition enables detection of the neurophysiological changes described by the haemodynamic response function, recorded during an fMRI scan. The spatial resolution of the EPI scan is however reduced due to the high speed of acquisition.

2.4.2.1.1 Functional MRI Analysis

As with structural images, functional MRI volumes must also undergo a series of preprocessing steps, however, in comparison, these steps must account for the fact that images are acquired over time. The main steps in preprocessing fMRI data therefore are realignment, normalisation, and spatial smoothing.

2.4.2.1.1.1 Preprocessing

In parallel with VBM, for functional images to be analyzed they must be normalized to a standard template and smoothed using an isotropic kernel. Since each subject has multiple images however, these must first be realigned to account for intra-subject variation, prior to the images being normalized to a standard template which accounts for inter-subject variation.

During realignment, the images acquired during each experimental session are realigned separately, in order to correct for artifacts in signal intensity resulting from subject head motion during imaging acquisition. This involves the first image of each session being realigned to the first image of the first session. The remaining images are then realigned to the first image of their respective session. This is achieved by estimating the 6 parameters (reflecting three translations and three rotations about orthogonal axes) that are required to realign one image with another using rigid body affine transformation. These parameters are then applied by reslicing the respective images using sinc interpolation which brings the images into alignment. During this process, a mean image for each subject based on their realigned images is also created. Once realigned, in order to alter the overall size and shape of each brain image so that it is comparable to a given template and placed in a standard co-ordinate space to allow for group level comparison, the images must be spatially normalized (Friston et al., 1995). Normalizaton of fMRI data comprises the realigned images for each subject being warped, using nonlinear-basis functions, to a template conforming to an internationally standardized size and shape, which for fMRI is the echo-planar image template. Importantly, this template adheres to an internationally recognized standard anatomical space, which in SPM8 is defined with respect to the standard international consortium for brain mapping (ICBM) template, resulting in the images for each subject being in

the same standardized stereotaxic space allowing for inter-subject comparison. To achieve this, the mean functional image for each subject is used to estimate the warping parameters that map this specific image onto the specified template via a 12-parameter affine transformation facilitated by a set of non-linear basis functions to optimize smooth normalization. Finally, the realigned and normalized images are spatially smoothed using an isotropic Gaussian kernel. As with structural MRI, by selecting the appropriately sized kernel, regional alterations in functional activation can be averaged across subjects using a spatial scale at which homologous regions of functional anatomy exist across all subjects. Secondly, this is performed to compensate for residual variability in functional anatomy after spatial normalization allowing the application of GRF theory for adjusted statistical inference.

2.4.2.1.1.2 Standard Univariate Analysis

Also like VBM, standard analysis of functional images is performed on a voxelwise basis founded on the GLM. The overarching principle of fMRI design is to elicit a known functionally specific neurovascular response through manipulation of the subject's behavior or experience. Subsequent univariate analysis of functional data comprises estimation of task related activity at each voxel, represented by BOLD contrast measurement, and an attempt to explain it as the weighted sum of the effects of a series of experimental conditions. These effects are the neural activity, believed by the investigator to have been elicited by a specific trial, convolved with the HRF inclusive of an approximate five second delay to account for the known delay in neurovascular coupling. To achieve this, the analysis is divided into a 'first-' and 'second-level'.

In the 'first-level' of statistical analysis the parameter estimates, representing the relative contribution of each experimental condition to the activity in a given voxel, are

obtained for each single subject. This is achieved by brain responses observed during a given condition, which are hypothesized by the investigator to be the result of the mental activity associated with that condition, being encoded by a set of regressors defining those conditions. These are then entered into a multiple linear regression which uses a least squares fit to provide parameter estimates (beta values) describing each regressor's contribution to the observed response at each voxel, in addition to a residual error term. This latter term reflects variance within the observed time series that is unaccounted for by the hypothesized effects. Once obtained, a contrast of parameter estimates between conditions can be made to identify activity associated with one specific condition over another.

Statistical parametric maps (SPMs) can then be created for each subject using the t-statistic generated for each voxel through the contrast of conditions based on the size of the parameter weights relative to their error term. In a task comprising two experimental conditions for example, with the first involving finger tapping using the left hand, and the second the right hand, an SPM can be created showing the variance in brain activity that is specifically associated with either the first, or second task.

In order to investigate group-level differences, a 'second-level' analysis must be performed. Here, the parameter estimates generated during the first-level for each individual subject can be combined to perform a 'second-level' or random effects analysis. This involves one observation per subject per condition being entered into a random effect model (usually a contrast of parameter estimates from the first level analysis) in order to compare the effect size of the contrast across groups relative to the inter-subject variability in these contrasts. Fundamentally, this statistical model allows one to make inferences about the population from which the sample of subjects was drawn. Based on the statistical approach used, a group level SPM can then be generated

displaying areas that are differentially active during a given condition in one group relative to another. Of note, since up to 100,00 voxels can be individually considered in a mass-univariate fashion for fMRI, the image must be corrected for multiple comparisons. The fact that the data points are not independent of each other however means that a standard Bonferroni correction cannot be used as it would be too conservative an approach. GRF theory is therefore employed which can take into account the lack of independence between voxels stemming from their continuity within the original data, and achieves for this 'continuous' data what a Bonferroni correction can for discrete data.

2.4.2.2 Hayling Sentence Completion Task

A neuropsychological test originally designed by Burgess and colleagues (Burgess, 1996), the Hayling sentence completion task (HSCT) requires subjects to read a sentence stem with the last word missing, and then overtly verbalise a word that is either congruent, or incongruent – semantically and phonologically - with the preceding stem. To date, the task has been shown by previous neuroimaging studies to elicit robust activation in prefrontal and lateral temporal regions within the language network (Collette, et al., 2001; Nathaniel-James, 1997). The neural activity elicited by the HSCT is associated with the processes of language initiation and suppression required to give a congruent or incongruent response to a preceding sentence stem. The test has also been used to examine executive dysfunction in patients suffering chronic schizophrenia (Chan, et al., 2006a) in addition to being used to show alterations in the temporal lobe region in those with a high genetic vulnerability to the illness (Whalley, et al., 2004). In 2008, Allen et al. adapted the task for use in fMRI, and subsequently in 2010, used it to

demonstrate significant differences both in regional activation and connectivity of areas associated with the language network between subjects with an ARMS and HCs.

In the current study, the HSCT consisted of 80 sentence stems selected from those given by Bloom and Fischler (Bloom, 1980) and Arcuri et al. (Arcuri et al., 2001) comprising five, six or seven words. In accordance with Allen et al. each sentence was assigned to either a response initiation, or response suppression condition. In the former condition, subjects were required to complete the sentence with a congruent response (i.e. He spread the butter using a KNIFE), and in the latter a semantically, and phonetically, incongruent response (i.e. The boy went to an expensive GIRAFFE). Each sentence was matched for word length across the two congruency conditions ensuring equal numbers of 5, 6 & 7 word stems in each. For both conditions the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm) was used to match the critical word in each sentence for frequency, concreteness, length and imageability. In each congruency condition, forty sentence stems were arranged into blocks each containing 5 sentence stems. During the task, each stem was presented visually to the subject in the MRI scanner one at a time. To control for inter-subject reading speed each word in the sentence stem appeared one at a time at an interval of 500ms. Each word appeared from left to right and the complete sentence stem remained on the screen for a further 500-1500ms after the last word of the stem appeared providing a total presentation time of 4 seconds for each word stem. Participants were then cued to articulate their verbal response by the appearance of a question mark that remained on the screen for a further 4 seconds. During this time a response was made before the presentation of the first word of the next sentence stem. There was thus a total inter-stimulus interval of 8 seconds between the presentations of each sentence stem, with each 5 sentence stem block lasting 40 seconds. The experimental condition was also contrasted with a control

condition consisting of overt articulation of the word 'REST' presented visually every 4 seconds preceded by a fixation cross that also lasted for 4 seconds (see Figures 2.2a-b). No reading only condition was used since previous work reports that differences in activation between reading a sentence and the initiation condition are relatively small - presumably due to a high predictability of the word used to complete the stem in both cases (Collette, et al., 2001; Kircher et al., 2001).

Prior to scanning, every participant received offline training on the task, with 10 initiation and 10 suppression practise sentence stems not used in the task itself. In the suppression condition it was stressed that responses should not be semantically related to the preceding sentence stem (e.g. The Earth is shaped like a SQUARE) nor phonologically related to the word to be inhibited (e.g. It was a cold and windy LIGHT). Once inside the scanner participants listened to a standardised instruction script before the Initiation phase and then again before the Suppression phase of the task.

For the purposes of error rate (see section 3.2.3.2 for more detail) and response time analysis, each subject's overt verbal responses were recorded by audio software (Cool Edit Syntrillium software). A subject's response time was defined as the period between the presentation of the question mark and the onset of their verbal response, the latency of which was measured using a software based voice trigger. Prior to each functional run, a dummy acquisition was employed during which the average power spectrum of the scanner noise was calculated and used as a noise profile. This profile was then used during the task, to digitally filter the microphone input signal based on a band pass filtering of the highest amplitude frequencies within the noise profile as well as non-linear subtraction method. The Root Mean Square (RMS) value of 8ms epochs of the differential of the filtered signal was then calculated, with speech onset determined by

the moment the RMS value exceeded a pre-set threshold set at just above scanner noise with no voice component.

Figure 2.2. Graphic example of the fMRI version of the Hayling sentence completion task

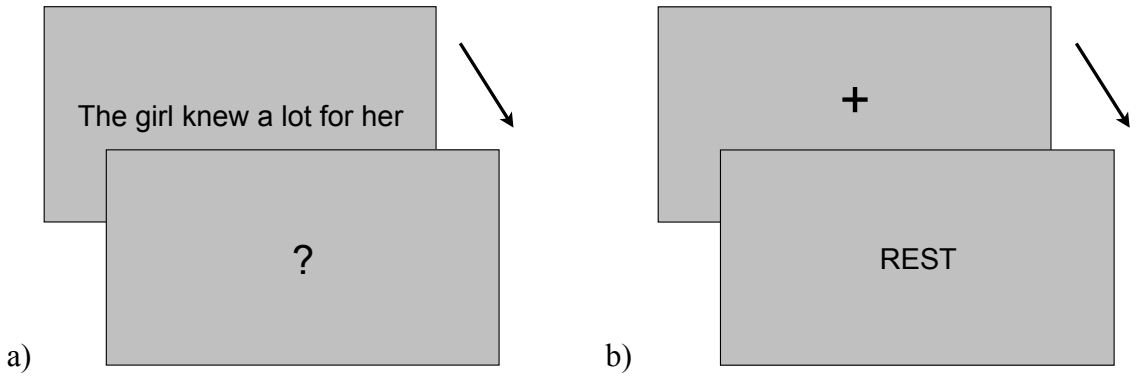


Fig. 2.2a) An example of a sentence stem valid for either the initiation or suppression condition. The image shows the complete sentence stem that would have developed over 4 seconds presented to the subject in the MRI scanner, followed by the second image of a '?' also displayed for 4 seconds, used to prompt an overt verbal response for the missing word (congruent for initiation, and incongruent for the suppression condition). Fig. 2.2b) An example of the rest condition, whereby the subject was presented with a visual fixation cross for 4 seconds, followed by presentation of the word 'REST' which they had to verbalise overtly a single time.

2.4.3 Diffusion Weighted Magnetic Resonance Imaging

Diffusion tenor imaging (DTI) is a form of neuroimaging that relies on the diffusion properties of water to create an image in three dimensions which principally visualises the microstructure of white matter tracts which connect inter- and intra-cerebral areas of the brain.

2.4.3.1 Diffusion

Based on the notion of Brownian motion, a diffusion-weighted image is one in which the relative orientation and ease of water diffusion within brain tissue is represented.

Brownian motion, or diffusion, describes the random movement of molecules within a fluid, each of which has an intrinsic diffusion coefficient governed by the viscosity of the medium it is in, its size and the ambient temperature of the environment. Within isotropic mediums water molecules are able to move freely in any direction impeded only by interactions with other water molecules, and over time, this movement follows a Gaussian displacement probability profile.

2.4.3.2 Anisotropy

If water molecules are impeded by a physical barrier and hence forced to travel in a given direction, then diffusivity of the molecule becomes direction dependent, a property referred to as anisotropy. This is the same scenario that occurs in white matter tracts. In contrast to the isotropic environment of a fluid, such as cerebro-spinal fluid (CSF), white matter tracts have numerous axonal membranes including the 'fatty' lipid-based myelin sheath for example. These act to impede the random free movement of water molecules, and as such, create an anisotropic diffusion gradient parallel to the tract itself (Moseley et al., 1990). Whilst the intrinsic diffusivity of the molecule remains the same, the anisotropic diffusion gradients are measured either perpendicular, or parallel, to the tract they are in. For a given WM tract, if the integrity of the surrounding membranes are good, and there are no obstructions in the tract itself impeding molecular movement, then the parallel anisotropy of the water molecules within the tract will be high and assigned a value of 1, on a scale of zero to one. In comparison, the perpendicular anisotropy of the water molecules, which are prevented by membranes acting as a barrier, will be close to or equal to zero.

2.4.3.3 Diffusion Tensor Imaging

In order to measure anisotropy of water molecules within the brain using MRI, diffusion weighted imaging exploits the fact that the MR signal is actually attenuated by the diffusive properties of water, with increased freedom of water movement within a voxel reflected by a greater reduction in signal. To achieve this, pulse sequences are implemented that contain additional gradients used specifically to encode the spatial location of protons over a set period of time from which the relative anisotropy value for each voxel can be assessed. A specific technique known as diffusion tensor imaging (DTI) builds on this concept by acquiring diffusion weighted data from at least six non-collinear directions (Basser & Pierpaoli, 1996). Using this data one can estimate a 3x3 symmetric matrix, describing the diffusion of a given anisotropic system along the x, y and z planes independent of the laboratory frame of reference, known as a diffusion tensor. Each tensor can be described as an ellipsoid defined by three orthogonal eigenvectors ($\varepsilon 1$, $\varepsilon 2$ and $\varepsilon 3$) representing the three principle Cartesian axes, each of which has an associated eigenvalue ($\lambda 1$, $\lambda 2$ and $\lambda 3$) indicating the relative diffusivity along that axis (see Fig. 2.3). In the case of neuroimaging, the principle eigenvector is considered to reflect the longitudinal direction of the underlying dominant white matter fibre.

Figure 2.3. Figurative example of a diffusion tensor using an ellipsoid shape

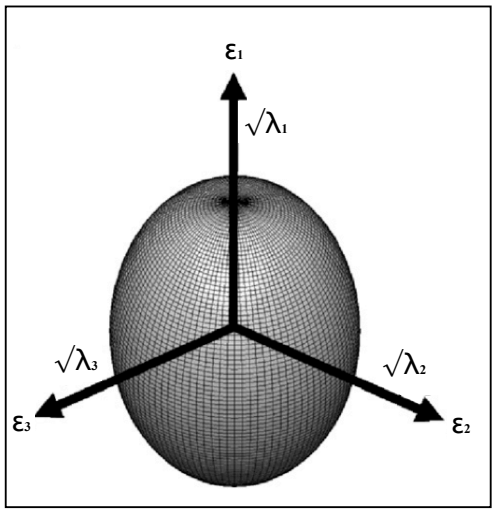


Figure 2.3 A figurative example of a diffusion tensor using an ellipsoid shape to demonstrate diffusivity, whereby the eigenvectors (ε1, ε2 and ε3) are scaled with respect to the square root of the eigenvalue (λ1, λ2 and λ3), and the shape itself represents the space to which a water molecule placed at its centre will diffuse with equal probability. Specifically, this tensor is representative of a molecule in which diffusion occurs principally along ε1 since λ1 > λ2 = λ3 such as would be expected in a WM tract. (Adapted from Jones, 2008)

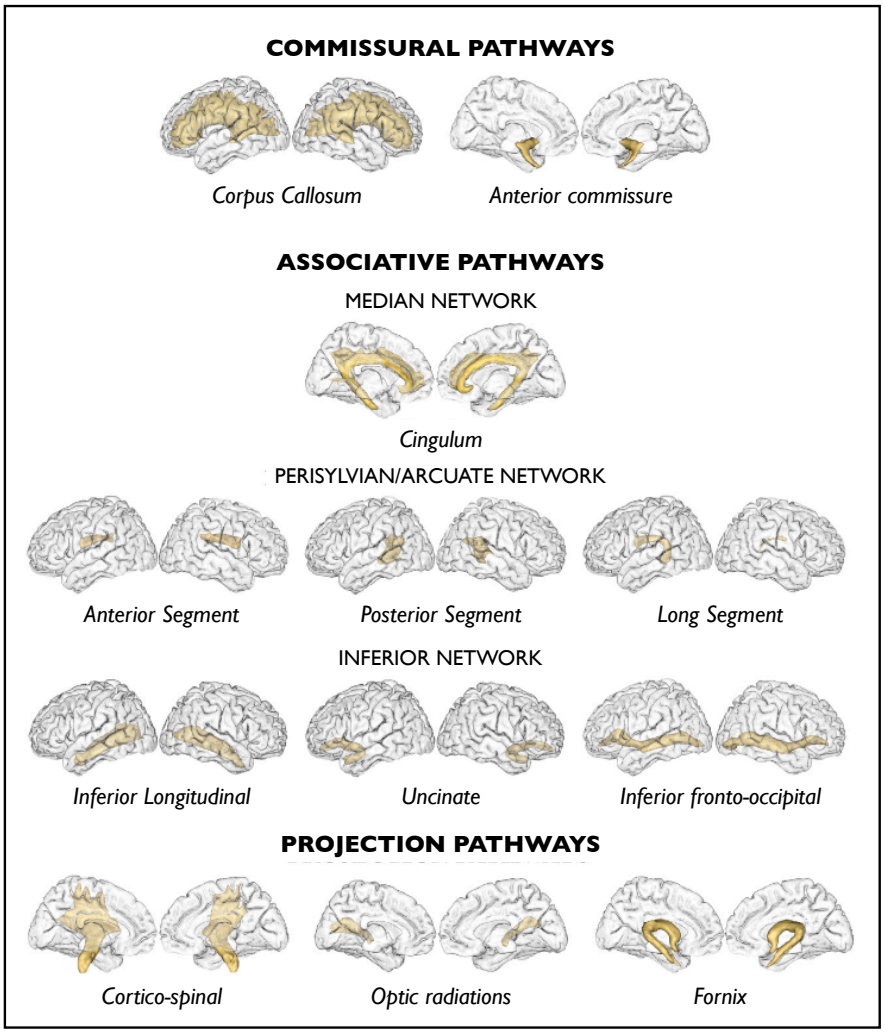
2.4.3.4 Fractional Anisotropy

Utilising the diffusion tensor a number of diffusion indices can be derived, one of which is known as fractional anisotropy (FA). Widely used in research, FA specifically refers to the fraction of the diffusion tensor that can be assigned to anisotropic, as opposed to isotropic, diffusion within each voxel and is characterised by the equation outlined below in which the mean eigenvalue, $\bar{\lambda}$ = (λ1+ λ2+ λ3) / 3;

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}}{\sqrt{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$$

This value is normalised to a scale of zero to one, by virtue of the multiplication constant, $\sqrt{(3/2)}$, with zero indicating isotropic diffusion and one representing diffusion constrained along a single axis (Jones, 2008). Based on this measure, DTI images can be generated in which increased FA values within a given voxel, representing increased anisotropic diffusion within that voxel, are visualised by increased voxel intensities, such that WM shows brightly on the image, whilst GM (which though anisotropic at the microscopic level is isotropic when averaged on the scale of a voxel (Pierpaoli et al., 1996)) and CSF show up as dark in colour, with those WM tracts containing the largest number of parallel running fibres (e.g. the corpus callosum) appearing brightest. Using this technique, it has been possible to virtually dissect, and subsequently visualise, the three major types of WM fibre that comprise the overall WM network in the human brain, namely association fibres, commissural fibres and projection fibres (see Fig. 2.4)

Figure 2.4. Visualisation of the three primary categories of WM tract (adapted from (Thiebaut de Schotten et al., 2011))



2.4.3.5 Diffusion Tensor Imaging Analysis

2.4.3.5.1 Preprocessing

Using whole brain FA images, it is also possible to generate a white matter (WM) 'skeleton' for each subject indicating the major WM pathways running through the brain (similar to those in Fig. 2.4). This can be achieved using the software package Tract Based Spatial Statistics (TBSS) (Smith et al., 2006). Specifically, TBSS is designed to prepare DTI derived images for an FA based voxelwise statistical analysis of WM, that parallels the GM based voxelwise statistical analysis implemented by

VBM. Overall, as outlined by the TBSS technical report TR05SS1 published by Smith et al. in 2006, the analysis process comprises three main steps; 1) identification of a common registration target to which the wholebrain FA image of each subject is aligned using nonlinear registration, 2) creation of a mean FA image based on the realigned images, followed by a 'thinning' of WM areas in order to generate a skeletonised FA image depicting the centres of each WM tract, and 3) projection of each individual's FA image onto the mean FA skeleton, with each voxel of the skeleton being filled with FA values corresponding to the nearest relevant WM tract centre for that subject, achieved by a search for maximum FA values that lie perpendicular to the local skeleton structure.

First, wholebrain FA images for each subject are partially eroded and zero end slices removed to account for likely outliers resulting from the diffusion tensor fitting. The nonlinear registration tool FNIRT (FMRIB non-linear image registration tool) (Andersson et al., 2007a, 2007b) is then used to register the images to a selected template and then affined transformed into a $1x1x1mm^3$ standard MNI152 space. The type of non-linear registration employed uses intermediate degrees of freedom to ensure that whilst the images are aligned sufficiently to allow local comparison between subjects, they are not warped to such a high degree that the fundamental structural topology of the image changes – as may be the case with 'perfectly' aligned images. These realigned, transformed images are then merged into a single four dimensional file. Using this 4D file a mean FA image based on all subjects is created and subsequently 'thinned' in order to create a mean FA skeleton, generated to represent the 'centre' of each white matter tract that is common to all subjects. This is achieved by non-maximum suppression guided by the voxel with the highest FA value perpendicular with respect to the local tract which is treated as the 'centre'. The wholebrain FA map for each subject is then projected onto this mean skeleton with the highest FA value at

each perpendicular point of the skeleton for that specific subject assigned to each corresponding skeleton voxel. This creates a subject-specific FA skeleton representing the major white tracts, in the same standard MNI space as every other subject.

2.4.3.5.2 Standard Univariate Analysis

Using the GLM graphical user interface (gui) in conjunction with the permutation based non-parametric 'randomise' programme, both contained within the FSL (fMRI of the Brain software library) toolbox, FA skeletons can be entered into a mass univariate standard voxelwise statistical analysis to identify differences at the group level.

Using the GLM gui, a design matrix is first created indicating the scans to be entered, corresponding regressors and the between groups contrast(s) of interest. This design matrix is then entered into the randomise programme which performs a permutation based non-parametric comparison between groups. Threshold free cluster enhancement can also be used which enhances cluster like structures in the data without changing the fundamental voxelwise property of the image. This results in an image, similar to a SPM, in which each voxel of the skeleton is designated a t-statistic indicating the level of difference in that voxel between groups. To correct for multiple comparisons, the image can be thresholded based on a permutation test, using a family wise error correction with $p<0.05$, such that the chance of one false positive occurring in the image is no more than 5% (Smith, et al., 2006).

2.5 Support Vector Machine

Machine learning (ML) is a field of artificial intelligence that has the overarching objective of creating automated techniques and algorithms able to extract information from different sources of data (Hastie, 2001). In comparison to the mass-univariate techniques discussed that generate group level inferences, supervised ML is a multivariate approach that allows characterisation at the level of the individual generating results which have a potentially substantive capacity for clinical translation. Specifically, supervised ML aims to discover regularities in data that can be used to classify the data into pre-defined categories, by developing a decision function, or classifier, that best captures the relationship between each 'example' of the data and the respective category, or label, to which it belongs (Burges, 1998). SVM is one type of supervised machine learning that aims to classify data points using an algorithm that maximises the margin between classes in a high dimensional space (Schölkopf & Smola, 2002). This algorithm, or function, operates by using an iterative procedure to map two or more operator defined categories, with two or more groups of observations by gradually reducing the difference between the predicted and expected result. Once optimised, the algorithm can be used to assign new, previously unseen, data to one of the predefined categories with a given accuracy (see section 5.2.1 for more detail). The processing pipeline for SVM can be described in three main steps. Firstly, the data that upon which the classifier is to be based must be prepared. Secondly, the classifier undergoes a period of training and testing and, thirdly, the classifiers performance is evaluated.

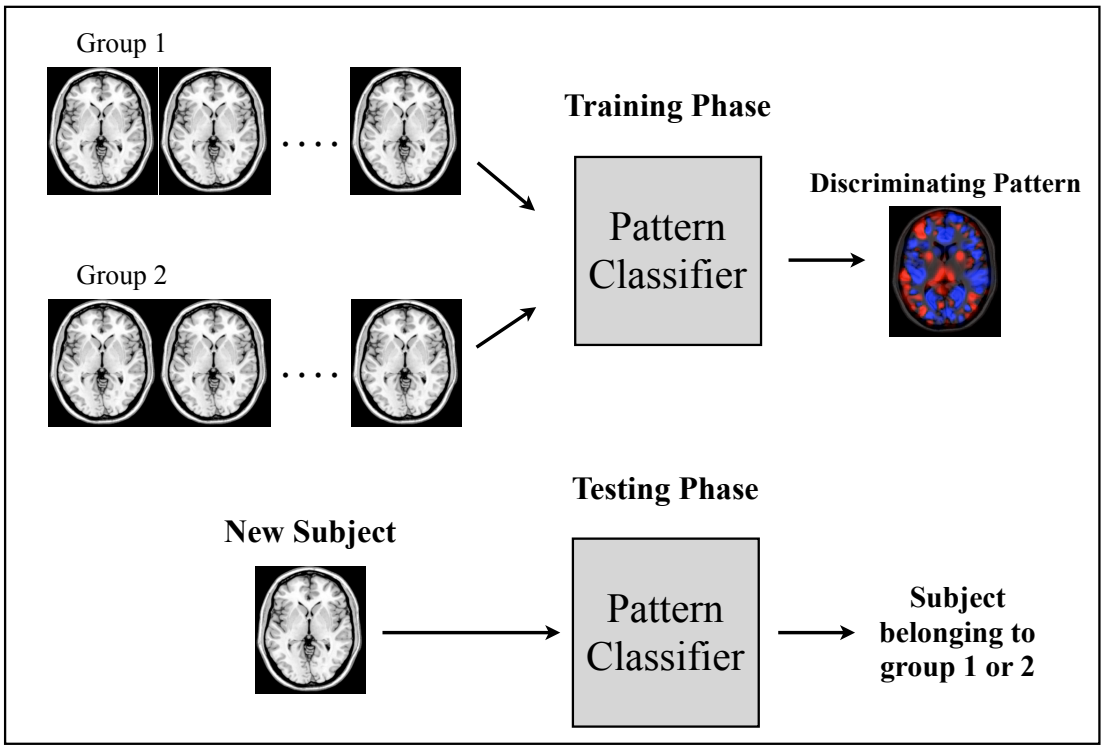
2.5.1 Preparation of the data for classifier training

Before the classifier is generated, the data to be used must first undergo a procedure known as 'feature extraction'. This involves transforming the data set from its original state into a set of features that can then be used as input into a SVM. This requires the entirety of the data to be encoded using a single column vector with each row representing a given feature. In the case of neuroimaging for example, each feature of a single column represents the intensity of a single corresponding voxel within one individual, with the correspondence between each specific voxel and each specific feature of the column remaining identical for every other subject. A second, optional, step that may also be performed is 'feature selection', used to select a subset of features from the data in order to i) reduce the dimensionality of the input feature space (e.g. number of voxels), and/or ii) facilitate learning by removing features that are effectively redundant for the purposes of classification. This can be achieved based either on *a priori* knowledge or the more data driven approach recursive feature elimination. Since the SVM employed here finds a solution for the OSH in kernel rather than feature space however, feature selection was not required to reduce dimensionality of the data, and will not be discussed further.

For the purposes of SVM, the length of each column vector for each subject must be identical, with each row of the vector corresponding to the same data element for every subject. This is to ensure that a pattern identified in the data able to differentiate between groups at the individual level refers to the same data elements in every subject, since the classification relies on the total amount of input data upon which the pattern is based. With respect to neuroimaging for example this means that the vector for each subject must contain the same number of voxels, in the same order. For genetic data

similarly, the SVM vector for each subject must pertain to the same SNPs, arranged in the same order.

Figure 2.5. Flowchart depicting the training and testing required for SVM classification using neuroimaging data as an example (adapted from (Orrù, et al., 2012))



2.5.2 Training and Testing

During the training and testing phase, a decision function is developed and optimised using a leave-one-out cross validation (LOOCV) framework (see Fig. 2.5). During the training phase the vector for each subject is treated as a data point in high dimensional space transformed through a linear transformation, using a linear kernel matrix. In this feature space, a linear decision function, or separating hyperplane, able to successfully distinguish the two groups is identified. Since there exists more than one way to separate the groups, the OSH is achieved by creating a hyperplane using those subject data points that are most difficult to classify, referred to as support vectors (Schölkopf & Smola, 2002), whilst maximising the distance between the two groups (see Fig. 2.6a-

b). The concept of maximising the distance between support vectors is motivated by the idea that this will increase the sensitivity of the subsequent OSH enabling it to accurately classify new unseen data points. Each data point in high dimensional feature space thus has a weight vector assigned to it reflecting its distance from the OSH and therefore its relative contribution in defining it. Hence, embedded in a LOOCV framework, an OSH is generated whereby all except one pair of data points (one subject from each group) are used as training data for the classifier and the remaining pair withheld as test data. The developed algorithm is then tested on the pair that has been omitted, with each 'test' subject classified as either class 1, or class 2, and the entire process repeated until every pair has been withheld once.

Figure 2.6

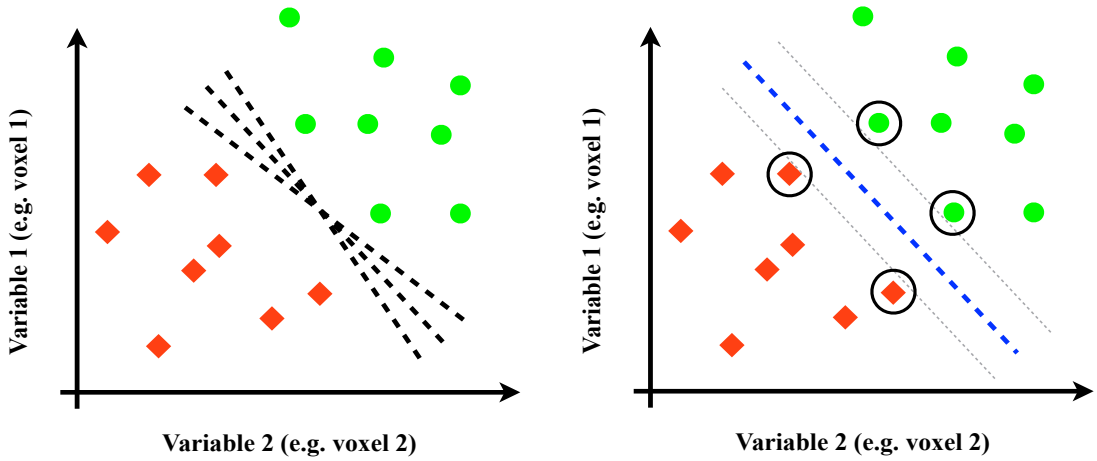


Figure 2.6: Pictorial representation of a pattern classification example involving two voxels. Each red diamond represents a subject in 'Group/Class 1' whilst each green circle represents a subject in 'Group/Class 2' in high dimensional feature space. For simplicity, we use a 2D example here. As Fig. 2.6a shows, initially there are a number of possible planes able to differentiate the data. Fig. 2.6b demonstrates application of margins (dashed grey lines), defined by the support vectors (circled diamonds and circles), which maximise the separation between groups defining the optimal separating hyperplane (dotted blue line) (adapted from (Orrù, et al., 2012)).

2.5.3 Classifier Performance

In order to quantify the performance of the SVM classifier, the mean of the proportion of true positive subjects correctly classified (sensitivity) and of the proportion of true negative subjects correctly classified (specificity) is calculated to provide an accuracy measure across all LOOCV folds (Hastie, 2001) (see also Box 1 in Chapter 4). Furthermore, in order to ensure a balanced accuracy using the LOOCV framework, the PROBID software package (www.brainmap.co.uk/probid.htm) employed here requires that individual subjects from each group are paired, matched here for age and gender, so that the classifier is unbiased with respect to either group (Lemm et al., 2011). An estimate of the statistical significance of the classifier's accuracy is obtained using a permutation test, whereby subjects are randomly allocated to a given class and the LOOCV cycle, as described above, repeated. In repeating this process one thousand times, a distribution of accuracies reflecting the null hypothesis that the classifier did not exceed chance is generated. Counting the occasions where the randomly permuted accuracy was equal to or greater than the true accuracy and dividing by 1000 then provides an estimated p-value for the significance of the classification algorithm.

2.5.4 Visualising SVM multivariate maps

Due to the linearity of the transformation used to convert data from input to feature space, it is possible to directly map the weight vector score assigned to each data point in feature space back into input space. This means that it is possible to generate a visual discrimination map displaying the pattern of features (e.g. voxels) able to differentiate between subject groups, with each feature represented by its weight vector score, i.e. the relative contribution made by that feature to the generation of the OSH (Mourão-Miranda et al., 2005). Due to the multivariate nature of the analysis, however, it should

be noted that such maps will explicitly contain the effects of interactions between features. In the context of neuroimaging, for example, this means that a voxel may be important in the differentiation of two groups, as reflected by a relatively large weight-vector score, either because there is a large difference in function, or structure, at the point, or because there is a small difference that is highly correlated with those in many other brain regions, with the voxel gaining importance from these correlations (Brammer, 2009).

In some previous studies (Marquand, et al., 2008; Mourão-Miranda, et al., 2005) a map threshold has also been applied to show only those features with a weight vector that exceeds a predefined threshold. In the current study no such threshold was applied, however, since any discrimination found was based on the total number of features (e.g. voxel intensities) entered into the SVM, and, since redundant feature extraction was not employed, nor *a priori* regions of interest specified in the context of neuroimaging data, unlike mass univariate results it is not possible to draw inferences regarding specific features (e.g. brain regions) of the discrimination pattern out of the context of the whole (Brammer, 2009).

2.5.5 Diagnostic classification using single modalities

After preprocessing, each subject's data (segmented GM images, FA skeletons, HSCT contrast images, orthogonally coded genotype data vectors and CVLT score vectors) were separately entered into SVMs (Burges, 1998) as implemented in the PROBID software package (www.brainmap.co.uk/probid.htm) running under Matlab 7.1 (Math Works, Natick, MA, USA). This was performed in order to assess the diagnostic potential for each modality with respect to ARMS and FEP subjects relative to HCs, and also to each other. For each comparison, the subject pairs matched for age (±4

years) and gender were used to construct samples for the classifier, with each individual scan, or data vector, treated as a data point located in high dimensional space and assigned by the operator to a given class; FEP, ARMS or HC. Using the LOOCV framework, classifiers were trained using all except one pair of subjects, upon whom the subsequent classifier was then tested. This process was then repeated, and the accuracy of the classifier derived from the mean of its sensitivity and specificity across all LOOCV folds. A permutation test was then performed with 1000 permutations to provide an estimate of the statistical significance of the accuracy.

2.5.6 Diagnostic classification using integrated modalities

In order to generate a single SVM classification decision based on multiple data types combined, there are currently a number of different techniques one can use (Kittler, et al., 1998). Broadly speaking, such techniques rely on one of two approaches, i) a single integrated kernel, single integrated classifier approach, or ii) a multiple kernel, multiple classifier ensemble approach. Of these two approaches, the first involves the creation of a single kernel matrix representing the kernel matrices of all base (i.e. single modality) classifiers being integrated. This integrated kernel is then used to train a single SVM classifier, meaning that the weight vectors for each data type are estimated jointly. In comparison, the second approach is underlined by a max-win strategy, whereby the final class label predicted for a given subject is determined by finding the class obtaining the largest number of predictions amongst the base classifiers. For this method, the weight vectors for each data type are therefore estimated independently.

In the current thesis, four distinct integrative methods were used and their results empirically compared. These were, 1) a ‘simple’ sum of kernels (SK), 2) multi-kernel learning (MKL), 3) prediction averaging (AV) and 4) majority voting (MV). Of these

four SK and MKL are examples of the first approach detailed above, whilst AV and MV are examples of the second.

In brief, whilst both SK and MKL involve the linear combination of kernel matrices to create a single integrated kernel matrix, MKL also incorporates an additional learning phase during which each base classifier is weighted in proportion to its relative contribution to the final decision function. With respect to AV and MV in comparison, although both rely on an ensemble approach to provide a single classification from multiple single classifiers, the decision reached by AV is based on the average prediction made for each subject by all classifiers, whereas only the binary decision of each classifier (i.e. class 1 or class 2) is considered in MV (see Fig. 5.1 in chapter 5 for a diagrammatic pipeline of each method).

For a more full and detailed description of each of these integrative methods please see chapter 5.

Chapter 3

Univariate Analysis of Neuroanatomical, Neurofunctional and Cognitive data, comparing First Episode Psychosis, At-Risk Mental State and Healthy Control Subjects

3.1 Introduction

Many studies to date have reported significant biological, and cognitive, differences in chronic schizophrenia (ChSz) patients relative to healthy controls (HCs) at the group level. Such reports include widespread changes to neuroanatomy (Ellison-Wright & Bullmore, 2009; Honea et al., 2005), neurofunction (Kindermann, et al., 1997; Minzenberg et al., 2009) and neuropsychological performance (Fioravanti et al., 2005; Heinrichs & Zakzanis, 1998) that are generally characterised by reductions rather than increases. More recently, similar, albeit less severe, alterations have also been reported in those in the earliest stages of psychosis, i.e. those with a recent first episode of psychosis (FEP) and those with a clinically increased risk of becoming psychotic, referred to as the at-risk mental state (ARMS) (Fusar-Poli et al., 2011a; Fusar-Poli et al., 2007b; Mesholam-Gately, et al., 2009; Smieskova et al., 2010). Based on univariate approaches however, the reports of such studies have been limited to group level inference only. Building on these results therefore, the main focus of the current thesis was to use multivariate machine learning analysis to assess the ability of different data types to successfully classify FEP and ARMS subjects at the single subject, rather than

group, level. However, to allow an anecdotal comparison between the outcomes of the two approaches, univariate analyses of the same data were also performed. These univariate analyses are reported in detail in the present chapter.

3.1.1 Previous studies of FEP and the ARMS

3.1.1.1 Structural MRI – Grey Matter

As highlighted in Chapter 1, magnetic resonance imaging (MRI) investigations of neuroanatomy may, broadly speaking, be divided into those examining grey matter (GM), and those examining white matter (WM). With respect to first of these, a small but growing number of studies suggest alterations similar to those seen in ChSz are evident in FEP and ARMS subjects, albeit to a less severe degree (Egerton, et al., 2011; Fusar-Poli, et al., 2011a). Frequently, such studies have employed an automated computerised technique that allows voxelwise analysis of structural brain images, known as voxel based morphometry (VBM) (Ashburner & Friston, 2000). Without the requirement of *a priori* defined regions of interest, VBM allows for a relatively explorative and unbiased approach for identifying areas of group level difference between groups of interest. In 2008 Borgwardt et al. for example used VBM to perform a within-subject prospective study to investigate alterations in GM before and after psychosis onset. Based on a cohort of 20 ARMS subjects categorised using the BSIP, of which half had developed frank psychosis at three year follow up, the group reported psychosis onset to be associated with GM reductions in frontal, temporal and parietal cortical regions. In comparison, they found no longitudinal differences in ARMS subjects who did not transition (Borgwardt, et al., 2008). These results are consistent with reports by Witthaus et al. who performed a cross sectional comparison of 30 ARMS categorised using the SIPS, 23 FEP and 29 HC subjects based on VBM analysis.

Relative to HCs, they reported that ARMS subjects showed significant GM reductions in the right inferior and superior temporal gyri, and hippocampus bilaterally. In comparison, FEP subjects relative to HCs showed significant GM reductions in the right anterior cingulate gyrus, bilateral middle and posterior cingulate gyri, right superior temporal gyrus and right hippocampus. Moreover, when compared directly, FEP subjects showed significant GM reductions in bilateral cingulate cortices, left orbitofrontal gyrus, right inferior frontal and superior temporal gyri, bilateral hippocampi, left parahippocampus, left amygdala, and left fusiform gyrus relative to ARMS subjects (Witthaus, et al., 2009). In another study recently published by Mechelli et al., multi-centre data was combined yielding 167 HCs and 182 ARMS subjects, of whom 48 ARMS subjects developed psychosis during two years of clinical follow up. Based on these cohorts, they reported the ARMS group as a whole to be associated with significant GM reductions of frontal regions bilaterally, whilst in comparison, ARMS subjects who went on to develop psychosis relative to those who did not had significantly less GM volume at baseline in the left parahippocampal cortex (Mechelli, et al., 2011). In comparison to these relatively consistent studies however, some studies have also reported finding no significant GM differences in those with an ARMS. For example, using a region of interest, rather than VBM, analysis Velakoulis et al. reported significant volume reduction in ChSz and FEP subjects in the bilateral and left hippocampi respectively, but found no evidence of baseline hippocampal change in subjects with an ARMS whether they subsequently converted to psychosis or not (Velakoulis et al., 2006). However, in accordance with a recent meta-analysis by Fusar-Poli et al., when taken together studies investigating GM in those with an ARMS and those with a FEP are largely consistent. Based on nineteen studies comprising 701 HCs, and 896 high-risk subjects (comprising both clinical and genetic high-risk), the

authors report that in the clinical high-risk group specifically, significant GM reductions were evident in the left hippocampus, insula, right superior temporal gyrus and right prefrontal cortex relative to HCs. In addition to this, they found significant GM reductions were also evident in the right inferior frontal and superior temporal gyri of those clinical high-risk subjects who went on to develop frank psychosis compared to those who did not (Fusar-Poli, et al., 2011a). Overall therefore, these results are consistent with the notion that GM alterations qualitatively similar to those seen in ChSz are also evident in ARMS and FEP subjects and become increasingly pronounced with progression from the former to the latter (see section 1.3.2.1 for further detail).

3.1.1.2 Diffusion Tensor Imaging – White Matter

With respect to investigations of WM alteration, a comparably small but growing body of literature is currently available that suggests changes qualitatively similar to those observed in ChSz are also evident in FEP and ARMS subjects albeit to a less severe degree (Pettersson-Yeo, et al., 2011). Of these, many such studies report a diffusion tensor imaging (DTI) derived measure of WM integrity called fractional anisotropy (FA) - in brief, FA is thought to represent an index of WM integrity within a given voxel, or tract, based on the diffusion properties of water molecules in that region, or regions (Jones et al., 1999) (for further detail see section 2.4.3.5). In a recent study by Carletti et al. for example, a comparison of DTI derived FA measures was made between 15 FEP, 32 HC and 32 ARMS subjects, of which 8 subjects from the ARMS group subsequently developed frank psychosis. With respect to FEP subjects firstly, significant FA reductions were reported in many different regions including the left inferior longitudinal and fronto-occipital fasciculi, posterior thalamic radiation and superior longitudinal fasciculi, in addition to the internal capsule, posterior and superior

corona radiata bilaterally, and body and splenium of the corpus callosum. Secondly, in ARMS subjects, the authors reported significant FA reductions in left inferior and superior longitudinal and fronto-occipital fasciculi, left posterior thalamic radiation, and also the splenium and body of the corpus callosum. Furthermore, they report that whilst qualitatively similar, in terms of scale and severity, alterations observed in the ARMS group appear to lie intermediate relative to the FEP and HC groups. Consistent with the findings from GM studies, the team also reported that ARMS subjects who developed frank psychosis were associated with additional longitudinal reductions in FA relative to those who did not, in this case localised to the left frontal lobe (Carletti, et al., 2012). Inconsistent with these results however, there have also been some studies published which reportedly found no significant FA difference, either between FEP and HC subjects (Friedman, et al., 2008), or between ARMS and HC subjects (Peters, et al., 2008). Nevertheless, when combined, the findings by Carletti et al. are in overall agreement with a recent review by Pettersson-Yeo et al., which examined DTI studies of ChSz, FEP and ARMS subjects performed to date. Specifically, from the limited literature available, the review's findings suggest that FEP and the ARMS appear to be associated with widespread alterations of WM integrity, primarily reductions, that frequently implicate regions of the frontal lobe, and that are qualitatively similar to those observed in ChSz albeit less severe (Pettersson-Yeo, et al., 2011) (see section 1.3.2.2 for further detail).

3.1.1.3 Neuropsychological Profile – California Verbal Learning Test

In concordance with investigations of neuroanatomy, many studies have also been published reporting deficits in multiple cognitive domains in FEP and ARMS subjects,

that appear similar to, but less severe than, those evident in ChSz, and which appear to get progressively worse with the transition from ARMS to FEP (Brewer, et al., 2005; Mesholam-Gately, et al., 2009; Wood, et al., 2007a). The California verbal learning test (CVLT) for example is an assessment aimed at quantifying an individual's capacity for verbal learning, retention and retrieval (Delis, et al., 1987), that has been used by a number of studies which report cognitive deficits in each of these domains in ChSz patients relative to HCs (Paulsen, et al., 1995; Tracy, et al., 2001). More recently, other groups have applied this test to subjects with a FEP, or ARMS, reporting similar findings as those seen in ChSz groups. Friis et al. for example reported that on average 219 FEP subjects performed significantly worse on the CVLT, amongst others, than would be expected based on demographically matched norms (Friis, et al., 2002). Consistent with these results a recent meta-analysis by Mesholam-Gately et al. also reported that immediate verbal memory, of which the CVLT is used as a measure, showed the largest effect size between 2204 FEP and 2775 HC subjects relative to other cognitive domains tested (Mesholam-Gately, et al., 2009). With respect to ARMS subjects in comparison, similar results have also been reported. For example, based on a comparison of 38 ARMS assessed using the SIPS, and 39 HC subjects, Lencz et al. reported significant verbal memory performance deficits in ARMS subjects relative to HCs of which the CVLT was used as a measure (Lencz, et al., 2006). Similarly, in a more recent longitudinal study by Niendam et al., the authors were able to give insight into the relationship between neurocognitive performance and functional ability. Specifically, based on a cohort of 35 ARMS subjects assessed using the SIPS, measured at baseline and at 8-month follow up, they reported a clear positive correlation, comparable to that observed in established psychosis, between verbal learning and

memory – of which the CVLT was used as a measure – and functional stability (Niendam, et al., 2007) (see section 1.2.3 for more detail).

3.1.1.4 Functional MRI – Hayling Sentence Completion Task

Based on neuropsychological tasks adapted for functional MRI (fMRI), a range of neurofunctional deficits have also been demonstrated in FEP and ARMS subjects relative to HCs (Egerton, et al., 2011; Fusar-Poli, et al., 2007b). The Hayling sentence completion task (HSCT) is one such example which has been used with ChSz, FEP and ARMS subjects to demonstrate significant differences in response initiation and inhibition relative to HCs, either in task performance and/or its associated neural correlates. First used as a functional paradigm by Whalley et al. in 2004, the team demonstrated that during speech initiation those with a genetic high-risk of psychosis had significantly less activation in medial prefrontal, thalamic and cerebellar regions relative to HCs (Whalley, et al., 2004), with a follow up connectivity analysis of the same data reporting abnormally high ipsalateral connectivity between left parietal and prefrontal cortices in the high risk group relative to HCs (Whalley, et al., 2005). Notably, however, this was based on a covert version of the task with subjects reporting their answers after the scan had ended. In comparison, a fully adapted overt version of the task was later developed in 2008 by Allen and colleagues (Allen, et al., 2008), who, in 2010, applied it to 15 ARMS and 15 HC subjects. Based on this investigation, the authors reported that ARMS subjects showed significantly greater activation across a wide range of language associated areas for both conditions, with a group by task interaction showing significantly greater activity in the anterior cingulate bilaterally during the suppression condition in the ARMS subjects relative to HCs (Allen, et al., 2010). Of note, neither Whalley et al. nor Allen et al. observed a significant

performance deficit in ARMS relative to HC subjects that corresponded with the observed change in neurofunction. In comparison, in two studies where the HSCT was conducted with ChSz subjects, both reported significant deactivations in regions of the default mode network in HCs relative to the patient group, which, in turn, were associated with significant performance deficits in the patient group (Grosselin, et al., 2010; Schneider, et al., 2011). In accordance with these findings, whilst no study has yet performed the fMRI version of the HSCT with FEP subjects, evidence is available from neuropsychological studies that is consistent with the findings in ChSz reported above. Based on 78 FEP and 60 matched HC subjects for example, Chan et al. reported that the FEP group made significantly more errors than HCs suggesting a deficit in their ability to initiate, or inhibit, a given response (Chan, et al., 2006b). Collectively therefore, the combined neurofunctional and behavioural HSCT results from ARMS, FEP and ChSz studies are at least consistent with the notion that the difference between the ARMS and established psychosis is associated with a change in the scale and severity of neurofunctional deficit, which although appears not to hinder task performance in those with an ARMS, may ultimately result in greater levels of error and slower reaction time in accordance with clinical decline (see section 1.3.2.3.2 for further detail).

3.1.2 Aims and Hypotheses

Based on these results, in the current study I aimed to examine whether differences in neuroanatomy, neurofunction and/or neuropsychological performance were evident in FEP and ARMS subjects relative to HCs, qualitatively similar to those seen in ChSz where they have been more frequently reported. I aimed to achieve this by using a VBM derived measure of grey matter, a DTI derived measure of white matter, a measure of

neuropsychological performance based on the CVLT-II, and a measure of neurofunction based on the HSCT. Given the literature available, I hypothesised that at the group level, significant differences would be observable between:

i) FEP and HC subjects with respect to: a) GM, with reductions evident in multiple cortical and subcortical frontal, temporal and parietal regions of FEP subjects relative to HCs (see section 1.2.2.1.2), b) WM, with reductions evident in multiple areas including the inferior, superior and fronto-occipital fasciculi, the posterior thalamic radiations, the coronae radiatae, the internal capsule, and the corpus callosum of FEP subjects relative to HCs (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in parietal and posterior cingulate regions of HCs relative to FEP subjects based on similar studies of ChSz patients (Schneider, et al., 2011), and, d) neuropsychological performance, with FEP subjects performing significantly worse for all subcomponents of the CVLT-II relative to HCs (see section 1.2.3.2).

ii) ARMS and HC subjects with respect to: a) GM, with reductions evident in cortical and subcortical frontal, temporal and parietal regions of the ARMS subjects relative to HCs (see section 1.2.2.1.2), b) WM, with reductions evident in the inferior, superior and fronto-occipital fasciculi, the posterior thalamic radiations, the coronae radiatae, the internal capsule and the corpus callosum of ARMS subjects relative to HCs (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in the right caudate and anterior cingulate gyri bilaterally of ARMS subjects relative to HCs (see section 1.2.2.3.2), and, d) neuropsychological performance, with

ARMS subjects performing significantly worse for all subcomponents of the CVLT-II relative to HCs (see section 1.2.3.2).

iii) FEP and ARMS subjects with respect to: a) GM, with reductions evident in cortical and subcortical frontal, temporal, parietal and cerebellar regions of FEP relative to ARMS subjects (see section 1.2.2.1.2), b) WM, with reductions evident in the inferior, superior and fronto-occipital fasciculi, external and internal capsules, posterior thalamic radiations, coronae radiatae and corpus callosum of FEP relative to ARMS subjects (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in the anterior cingulate and caudate, in addition to increases in regions of the default mode network, in FEP relative to ARMS subjects, based on previous results of ChSz patients relative to HCs, and ARMS subjects relative to HCs, respectively (see section 1.2.2.3.2), and, d) neuropsychological performance, with FEP subjects performing significantly worse than ARMS subjects for all subcomponents of the CVLT-II based on previous results showing progressive cognitive decline associated with conversion from ARMS to FEP (see section 1.2.3.2)

3.2 Materials and Methods

3.2.1 Participants

Participant demographics, premorbid IQ and clinical data are presented in Table 3.1.

3.2.1.1 First Episode Psychosis

Nineteen subjects were recruited through the South London and Maudsley National Health Service Trust (http://www.slam.nhs.uk). All had experienced a FEP within the past 24 months that met DSM-IV criteria for a schizophreniform psychosis (see sections 1.1, 2.1.1 and Table 3.1 for more detail).

3.2.1.2 At-Risk Mental State

Nineteen subjects were recruited from OASIS (Outreach and Support in Southeast London), a clinical service for young people at high-risk of developing psychosis (Broome, et al., 2005). Their clinical status was defined according to the PACE (Personal Assessment and Crisis Evaluation) criteria (Yung, et al., 1998) and their diagnosis confirmed using the CAARMS (Comprehensive Assessment of At-Risk Mental States) (Phillips, et al., 2000). In brief, individuals are classed as having an ARMS based on the presence of 1) Attenuated psychotic symptoms, 2) Brief Limited Intermittent Psychotic symptoms, or 3) Trait and state risk factors (e.g. a first degree relative with a psychiatric disorder or schizotypal personality disorder) combined with a significant decline in cognitive and social functioning over the past year (see sections 1.2, 2.1.2 and Table 3.1 for more detail).

3.2.1.3 Healthy Controls

Twenty-three healthy control subjects were recruited from the local area through advertising. No subjects met criteria for a DSM-IV-TR psychiatric disorder, fulfilled the PACE criteria for prodromal symptoms nor had a first-degree family history of psychiatric disorder (see sections 2.1.5 and 2.2.3 for more detail).

All subjects included in the study were 18-35 years old and spoke English as their first language. Exclusion criteria included a history of neurological disorder, evidence of substance abuse and/or dependence as defined by DSM-IV-TR criteria, prior head trauma resulting in loss of consciousness and/or hospitalisation, or any contraindications to exposure to a magnetic field (e.g. metal implants, or pregnancy) (see section 2.1.4 for more detail).

3.2.1.4 Subject Pairing

As discussed in section 2.5.3, for the purpose of SVM implemented through PROBID (http://www.brainmap.co.uk/probid.htm), balanced numbers of subject pairs, matched here for age and gender, are required for each diagnostic comparison. Using these pairing criteria, from our cohort of 19 FEP, 19 ARMS and 23 HCs, 19, 19 and 15 subject pairs matched for age (±4 years) and gender, were available for the FEP versus HC, ARMS versus HC and FEP versus ARMS diagnostic comparisons respectively. Critically, these pairings were maintained for the standard analysis of sMRI, DTI, fMRI and neuropsychological data outlined below, to allow a more direct comparison with the results of the multivariate analyses detailed in Chapter 4. This also allowed paired, rather than independent, t-tests to be performed, providing the added benefit of the enhanced sensitivity afforded by a paired t-test due to the intrinsic reduction in the standard error of the mean (Peers, 1996; Zimmerman, 1997). Analysis of the HSCT

behavioural data using a repeated measures ANOVA was an exception to this for consistency with the few relevant previous studies. (See Table 3.1 for a detailed characterisation of the subject pairs used).

Table 3.1. Demographic information for the subjects used for each of the three diagnostic SVM classification comparisons, showing the mean followed by the standard deviation (in brackets).

	ARMS versus HC			FEP versus HC			FEP versus ARMS		
	HC (n=19)	**ARMS (n=19)**	***Analysis***	**HC (n=19)**	**FEP (n=19)**	***Analysis***	**ARMS (n=15)**	**FEP (n=15)**	***Analysis***
Age (years)	23.31 (3.43)	22.42 (3.42)		24.89 (4.41)	24.37 (4.71)		23.20 (3.43)	23.27 (3.69)	
Gender	9M:10F	9M:10F		12M:7F	12M:7F		9M:6F	9M:6F	
Laterality	18R:1L	15R:4L		19R:0L	19R:0L		13R:2L	15R:0L	
Years of Education	14.63 (3.48)	11.79 (3.55)		14.89 (3.67)	13.58 (4.33)		14.00 (4.29)	11.87 (3.67)	
WRAT-R estimated premorbid IQ (SS)	107.58 (10.77)	103.16 (13.14)	$t = -1.22$ $p = 0.237$	108.53 (10.48)	102.74 (9.33)	$t = -1.68$ $p = 0.110$	104.87 (11.98)	103.80 (9.97)	$t = -0.25$ $p = 0.807$
Ethnicity	12W:4B:1A	9W:7B:1A		12W:2B:0A	2W:8B:4A		3W:5B:4A	7W:5B:0A	
PANSS total[a]	30.47 (0.90)	52.53 (9.28)	$t = 10.38$ $p < 0.001$	30.58 (1.46)	54.37 (15.13)	$t = 6.81$ $p < 0.001$	53.73 (9.11)	51.80 (12.46)	$t = -0.46$ $p = 0.655$
PANSS positive[a]	7.26 (0.81)	12.84 (3.67)	$t = 6.76$ $p < 0.001$	7.10 (0.46)	12.58 (3.96)	$t = 6.09$ $p < 0.001$	12.80 (3.65)	12.07 (3.08)	$t = -0.60$ $p = 0.556$
PANSS negative[a]	7.05 (0.23)	14.00 (4.08)	$t = 7.51$ $p < 0.001$	7.00 (0.00)	13.79 (5.26)	$t = 5.63$ $p < 0.001$	14.33 (4.05)	13.47 (5.05)	$t = -0.51$ $p = 0.618$
PANSS general[a]	16.16 (0.50)	25.68 (5.01)	$t = 8.10$ $p < 0.001$	16.47 (1.43)	28.00 (8.35)	$t = 5.90$ $p < 0.001$	26.60 (4.97)	26.27 (7.35)	$t = -0.14$ $p = 0.893$
Total Medication[b]		43115.63 (57681.35)			34652.39 (29540.47)		43115.62 (57681.35)	31291.57 (27178.41)	
Mean Medication/day[c]		126.01 (84.21)			215.42 (108.39)		126.09 (84.21)	211.70 (109.89)	

[a] Symptom profile recorded at time of scan; [b] Total Medication refers to the mean absolute amount of medication taken by that group in standardized mg units of Chlorpromazine ± 1SD; [c] Mean Medication/day is the average medication dosage taken by each subject during their period of treatment in standardized mg units of Chlorpromazine ± 1SD; M = males; F = females; R = Predominantly Right Handed; L = Predominantly Left Handed; WRAT-R (SS) = Wide Range Achievement Test Revised (Standardized Score); PANSS = Positive and Negative Syndrome Scale; Ethnic Groupings - W = White British, B = Black British, A = Asian British.

3.2.2 Data Acquisition

3.2.2.1 Estimate of Premorbid IQ

During the pre-scan assessment interview, the reading subtest of the revised wide ranging achievement test (WRAT-R) (see section 2.2.2.1 for detail) was administered to each subject by a trained researcher and their scores recorded. Designed to assess a subject's ability to recognise and name letters and words (Jastak & Wilkinson, 1984) this specific subtest has been shown to provide an accurate estimate of premorbid IQ in clinical populations suffering brain dysfunction (Johnstone, et al., 1996) and is provided with associated demographically corrected norms (see Table 3.1).

3.2.2.2 Clinical Symptom Profile

During the pre-scan assessment interview, the positive and negative syndrome scale (PANSS) (Kay, 1987) (see section 2.2.3 for detail) was administered to each subject by a trained researcher and the subject's scores recorded. These scores were used for FEP and ARMS subjects as a measure of psychopathology on the day of scanning, and for HCs as a confirmation they were not experiencing any attenuated, or frank, psychotic symptoms.

3.2.2.3 Magnetic Resonance Imaging

All neuroimaging was conducted using a 3T MRI scanner (Signa, LX-GE, Milwaukee, USA) at the Maudsley Hospital, London.

3.2.2.3.1 Structural MRI

Structural images were acquired using a volumetric 3D Spoiled Gradient Recall (SPGR) sequence comprising a repetition time (TR) of 7.04ms, an echo time (TE) of 2.82ms, a flip angle of 20°, a slice thickness of 1.1mm, an in-plane resolution of 1.09mm x 1.09mm, a field of view of 21cm^2, and a 256x256matrix. In total 196 coronal slices were produced.

3.2.2.3.2 Diffusion Tensor Imaging

DTI volumes were acquired using a multi-slice peripherally-gated doubly refocused spin-echo echo-planar imaging sequence that was optimised for precise measurement of the diffusion tensor in parenchyma, from 60 contiguous near-axial slice locations with a TE of 104.5ms, flip angle of 90°, slice thickness of 2.4mm, field of view of 30.7cm^2 and a 128x128 matrix. As a cardiac-gated acquisition the effective TR for each subject ranged between 15 and 20 RR intervals; RR is the interval between the two R phases in the electrocardiogram 'QRS' heart cycle. The maximum diffusion weighting gradient was 1300 s/mm^2 (defined by time taken for one a molecule to move 1mm^2). At each slice location 32 diffusion-weighted images in which gradient directions were uniformly distributed in space in accordance with the recommendations of Jones and colleagues (Jones et al., 2002) were acquired alongside 4 images with no diffusion gradients which were used for comparison.

3.2.2.3.3 Functional MRI

Functional images were acquired using a TR of 2000ms, a TE of 30ms, a flip angle of 70°, a slice thickness of 3mm, a field of view of 24cm^2 and a 64x64 matrix. In total, 38 axial slices in parallel to the anterior commissure–posterior commissure (AC-PC) line

were collected for each subject. During the acquisition of functional images, subjects performed the HSCT using an experimental protocol detailed in section 2.4.2.2 and described elsewhere (Allen, et al., 2008). In brief, subjects were visually presented for 4s with a 5, 6 or 7 word sentence-stem with the last word omitted. Presentation of a question mark then required them to overtly generate a word either congruent (initiation condition) or incongruent (suppression condition), with the preceding sentence. The task was arranged into eight blocks of 5 sentence-stems, with each block separated by a baseline condition whereby the subject was shown a visual fixation-cross for 4s, followed by the word 'REST' for 4s which they had to read overtly. During the task, each subject's responses were recorded for subsequent performance analysis. Prior to the scan, subjects were familiarised with an offline version of the task using 10 initiation and 10 suppression practise sentence stems not included in the subsequent online version (see section 2.4.2.2 for more detail). Overall one initiation and one suppression session were run separately, generating 600 image volumes in total.

3.2.2.4 Neuropsychological Profile

The CVLT-II, is the second edition of a language and memory task designed to quantify an individual's ability for verbal learning, retention and retrieval (Delis, et al., 1987) and is provided with associated demographically corrected norms (Delis, et al., 2000). For the present investigation, the test was administered to each subject by a trained researcher during the pre-scan assessment interview and the subject's scores recorded. These were subsequently inputted into the associated CVLT-II software package (Delis & Fridlund, 2000). This software provides a comprehensive summary of raw and demographically corrected standardised scores for the different task components (see section 2.2.2.2 and Table 3.3 for more detail).

3.2.3 Data Analysis

3.2.3.1 Demographic Data

Paired t-tests were conducted using the IBM SPSS 19 statistical software package (http://www-01.ibm.com/software/analytics/spss/products/statistics/) in order to compare premorbid IQ for all diagnostic comparisons, and also the PANSS scores for the FEP versus ARMS subject comparison. PANSS score tests were not carried out for those comparisons involving HCs since differences in psychopathological score for this group relative to either the FEP or ARMS group was implicit. Inferences were made at $p<0.05$.

3.2.3.2 HSCT: Behavioural Analysis

Based on digital recordings of their overt responses during the task, the mean proportion of errors and mean reaction times were calculated for each subject. In accordance with the Hayling and Brixton Test section 5 (Thames Valley Test Company Limited, 1997), errors in the initiation condition were defined as responses that were semantically incongruent with the preceding sentence stem, whilst errors in the suppression condition were defined as responses that were semantically or phonologically congruent with the preceding sentence stem. The appropriateness of each response was assessed first by a single investigator, who, if uncertain of their decision, would consult a second investigator. In the event a consensus decision could still not be reached, the opinion of a third investigator would then be sought for a casting vote. All assessors were native English speakers. In order to analyse mean proportion of errors and mean reaction time (RT) for each group a repeated measures ANOVA was performed using IBM SPSS 19 statistical software package (http://www-01.ibm.com/software/analytics/spss/products/statistics/), with congruency as a within

subject factor and group as a between-subject factor. Equal variances were assumed unless Levene's test of equality of variance was significant at 95% level. The groups comprised the total number of subjects from which the SVM pairs were generated, representing 19 FEP, 19 ARMS and 23 HC subjects. Of these, the behavioural results for one HC were omitted due to a microphone error during the scan that resulted in some responses not being recorded. Where there was a significant congruency-by-group interaction, a post-hoc univariate ANOVA pairwise comparison with Bonferroni correction was performed for the initiation and suppression conditions separately, with the proportion of errors made by each subject in the initiation, or suppression, condition as the dependent variable and group as the fixed-factor.

3.2.3.3 Structural MRI

Preprocessing was conducted using SPM8 software (http://www.fil.ion.ucl.ac.uk/spm) running under Matlab 7.1 (Math Works, Natick, MA, USA). First, all images were manually reoriented to place the anterior commissure at the origin of the three-dimensional Montreal Neurological Institute (MNI) coordinate system. The unified segmentation procedure (Ashburner & Friston, 2005) was then used in combination with the Diffeomorphic Anatomical Registration through the Exponentiated Lie algebra (DARTEL) algorithm (Ashburner, 2007) to segment and subsequently normalise the reoriented structural images. This is achieved first by the creation of a study specific template derived from the mean of all subjects' images. The deformations from this initial custom template to the image of each individual subject are then computed. The inverse of these deformations are then applied to each individual image, which are then used to generate a new custom template. This process is iteratively repeated a set number of times, and the final deformation parameters used to warp the image of each

subject with a high level of precision to the same standard (MNI) space. In contrast to using a pre-generated template, this technique allows each subject's image to be segmented with greater accuracy and sensitivity (Yassa & Stark, 2009). In the current study an additional "modulation" step was also implemented to account for the expansions and/or contractions of brain regions during the normalisation process that in turn alter the volumetric measurements for each voxel. This consisted of multiplying the spatially normalised grey matter after normalisation, by the volumetric changes encoded during the warping procedure implemented for normalisation, so as to ensure that the total amount of signal in each voxel was conserved. As a final preprocessing step, the resulting modulated images were smoothed by convolving with an 8mm isotropic kernel at full width half maximum (FWHM). GM images were analysed using a paired t-test implemented in SPM8 software (http://www.fil.ion.ucl.ac.uk/spm) running under Matlab 7.1 (Math Works, Natick, MA, USA), with statistical inferences made at $p<0.05$ after Family-Wise Error (FWE) correction for multiple comparisons (see section 2.4.1.6.2 for more detail).

3.2.3.4 Diffusion Tensor Imaging

Diffusion data was processed using the software package ExploreDTI (Leemans et al., 2009). Data was first pre-processed correcting for eddy current distortions and head motion. For each subject the b-matrix was also reoriented to provide a more accurate estimate of tensor orientations (Leemans & Jones, 2009). Diffusion tensors were estimated using RESTORE (Chang et al., 2005), an automatic and iterative approach for the automatic rejection of data outliers. Finally fractional anisotropy (FA), mean diffusivity (MD) and eigenvalue maps were then generated. FA skeletons were then created from the FA maps using the software package Tract-Based Spatial Statistics

(TBSS) (Smith, et al., 2006). To achieve this, the images were first partially eroded and zero end slices removed to account for likely outliers resulting from the diffusion tensor fitting. The nonlinear registration tool FNIRT (Andersson, et al., 2007a, 2007b) was then used to non-linearly register the images to the FMRIB58_FA template aligning them to a 1x1x1mm standard MNI space, and then merged into a single 4D file. From this file a mean FA image was created which was used to generate a mean FA 'skeleton' representing WM tracts common to all subjects. Each subjects FA image was then projected onto the mean FA skeleton with the highest FA value at each point of the skeleton for that subject assigned to each corresponding skeleton voxel. Each subject's skeleton is hence specific to them and also in the same standard (MNI) space as the rest of the group. For more in depth detail see Smith et al. (2006). Paired t-test analysis of FA skeletons was performed using TBSS, whereby the FA skeletons for each group were fed into voxelwise cross-subject statistics. Analysis identifying differences between groups was made using a voxelwise statistical analysis based on a nonparametric permutation test (5000 permutations) and with a general linear model design matrix. Results were obtained using the Threshold-Free Cluster Enhancement (TFCE) method (Smith & Nichols, 2009) and with a significance equal to a *p*-value of less than 0.05 FWE corrected for multiple comparisons.

3.2.3.5 Functional MRI

Functional images were pre-processed using SPM8 software (http://www.fil.ion.ucl.ac.uk/spm) running under Matlab 7.1 (Math Works, Natick, MA, USA). Initiation and suppression conditions were entered as two separate sessions, and a standard event-related design was conducted whereby the first image of the suppression condition was realigned to the first image of the initiation condition. The

remaining images were then realigned to the first image of their respective session and resliced with sinc interpolation. Finally, the images were spatially normalized (Friston, et al., 1995) to a standard MNI-305 template using nonlinear- basis functions and spatially smoothed with a 6-mm full width at half maximum isotropic Gaussian kernel, thus compensating for residual variability in functional anatomy after spatial normalization, permitting the application of Gaussian random field theory for adjusted statistical inference. A standard event-related first-level analysis of regional responses was performed to identify regional activations in each subject; this involved convolving the onset times (i.e. the onset of the question mark prompting a verbal response) with a canonical haemodynamic response function. To exclude low frequency drifts the data was high-pass filtered using a set of discrete cosine basis functions with a cutoff of 128sec. Six experimental conditions: 1) initiation 2) suppression 3/4) overt oral repetition of the word 'REST' during initiation/suppression and 5/6) cross-fixation during initiation/suppression were modeled independently. To remain consistent with the format of the data used in the multivariate analysis described in the next chapter, therefore allowing for a closer anecdotal comparison, error responses were not removed from the initiation or suppression conditions through the use of a separate regressor; however, this means that potential 'noise' in the data due to individual differences in behavioural performance was not removed which may in turn have reduced statistical sensitivity. Using the GLM parameter estimates obtained for all brain voxels five contrasts of interest were then computed, namely; Suppression>Initiation, Initiation>Repetition of 'REST' during Initiation, Suppression>Repetition of 'REST' during Suppression, Initiation>Cross-Fixation during Initiation, Suppression>Cross-Fixation during Suppression. Second level analysis of HSCT contrast images was performed using a paired t-test implemented in SPM8 software

(http://www.fil.ion.ucl.ac.uk/spm) running under Matlab 7.1 (Math Works, Natick, MA, USA), with statistical inferences made at $p<0.05$ after FWE correction for multiple comparisons, and cluster threshold $k\geq10$. (see section 2.4.2.1.1.2 for more detail regarding fMRI data analysis and the HSCT).

3.2.3.6 Neuropsychological Profile

CVLT-II scores were compared using a paired t-test implemented in IBM SPSS 19 statistical software package (http://www-01.ibm.com/software/analytics/spss/products/statistics/). A Bonferroni correction was applied to account for the seven separate test domains of the CVLT-II (see domains A-G in Table 3.3), with inferences of statistical significance therefore made at $p<0.007$ (i.e. $\alpha=0.05/7$). For completeness, uncorrected results are also reported but not discussed. Due to one ARMS subject not completing the task in full, the comparison of ARMS versus HC subjects was based on 18 subject pairs. Consistent with standard analysis of cognitive data only standardised scores were compared for measures where both raw and standardised scores were available.

3.3 Results

3.3.1 Demographics

Subjects in each group were paired according to age and gender. There were no significant differences with respect to premorbid-IQ between any of the groups (p>0.05), nor was there a significant difference in PANSS scores (total, positive, negative or general) between ARMS and FEP subjects (p>0.05). With respect to medication, all FEP subjects, except one, were medicated. In comparison all ARMS subjects were medication-naïve, apart from two (see Table 3.1).

3.3.2 HSCT Performance

Proportion of Errors: A repeated measures ANOVA showed a significant main effect of congruency (F = 41.761, df = 1, p = 0.000) due to a greater proportion of errors occurring in the suppression condition. The main effect of group was also significant (F = 6.519, df = 2, p = 0.003); a repeated measures ANOVA pairwise comparison with Bonferroni correction showed that ARMS (mean difference = 0.097, SEM = 0.029, p = 0.005) and FEP (mean difference = 0.082, SEM = 0.029, p = 0.021) subjects made a significantly higher proportion of errors than HCs. Mean proportion of errors did not differ significantly between FEP and ARMS subjects (mean difference = -0.015, SEM = 0.029, p = 1.000). There was also a congruency-by-group interaction (F = 3.423, df = 2, p = 0.039); a post-hoc one way ANOVA pairwise comparison with Bonferroni correction showed that in the suppression condition, ARMS (mean difference = 0.149, SEM = 0.044, p = 0.004), and FEP (mean difference = 0.117, SEM = 0.044, p = 0.032) subjects made a significantly higher proportion of errors than HCs. FEP and ARMS did

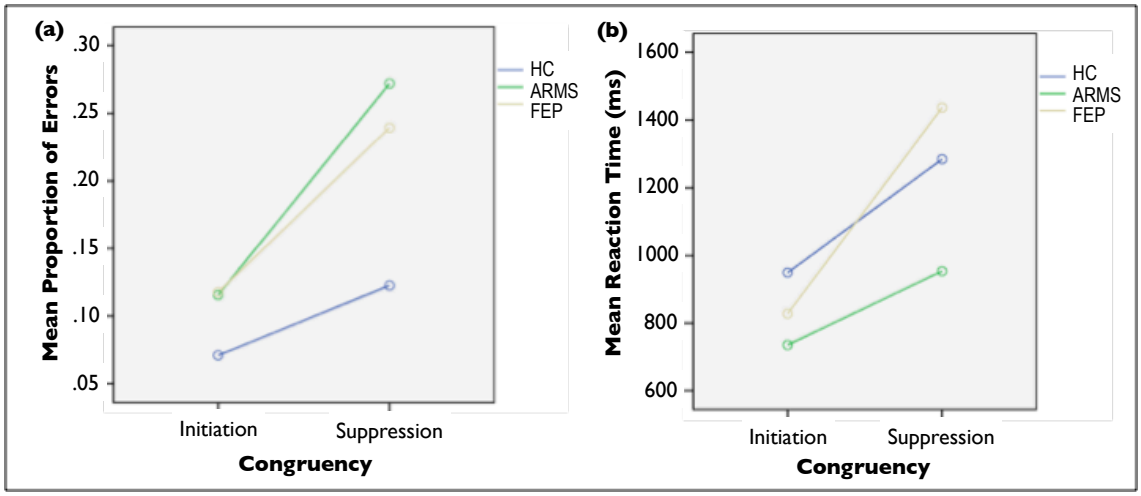
not significantly differ (mean difference = -0.033, SEM = 0.046, p = 1.000). The proportion of errors during the congruency conditions are shown in Fig. 3.1a and reported in detail in Table 3.2.

Reaction Times: A repeated measures ANOVA showed a significant effect of congruency (F = 54.966, df = 1, $p < 0.001$) due to significantly longer reaction times (RTs) on average in the suppression condition (see Fig. 3.1b and Table 3.2). The main effect of group was also significant (F = 3.766, df = 2, p = 0.029); however, a repeated measures ANOVA pairwise comparison with Bonferroni correction did not find a significant difference between ARMS and HC subjects (mean difference = -272.171, SEM = 114.983, p = 0.064), FEP and HC subjects (mean difference = 15.396, SEM = 114.983, p = 1.000), nor FEP and ARMS subjects (mean difference = 287.568, SEM = 119.115, p = 0.057). There was also a congruency-by-group interaction (F = 4.711, df = 2, p = 0.013); a post-hoc one-way ANOVA pairwise comparison with Bonferroni correction showed that in the suppression condition, on average FEP subjects had significantly longer RTs than ARMS subjects (mean difference = 482.690, SEM = 160.156, p = 0.012).

Table 3.2. Mean proportion of errors and reaction times for each group during the HSCT.

Condition	**Mean Proportion of Errors (± SD)**		
	HC	**ARMS**	**FEP**
Initiation	0.071 (0.055)	0.116 (0.099)	0.118 (0.078)
Suppression	0.123 (0.100)	0.272 (0.171)	0.239 (0.148)
Condition	**Mean Reaction Time (± SD) (ms)**		
	HC	**ARMS**	**FEP**
Initiation	949.74 (332.45)	736.043 (282.53)	828.67 (361.49)
Suppression	1284.47 (596.33)	953.82 (443.56)	1436.51 (400.00)

Figure 3.1. Graph showing (a) mean proportion of errors, and (b) mean reaction time for each group with respect to congruency.



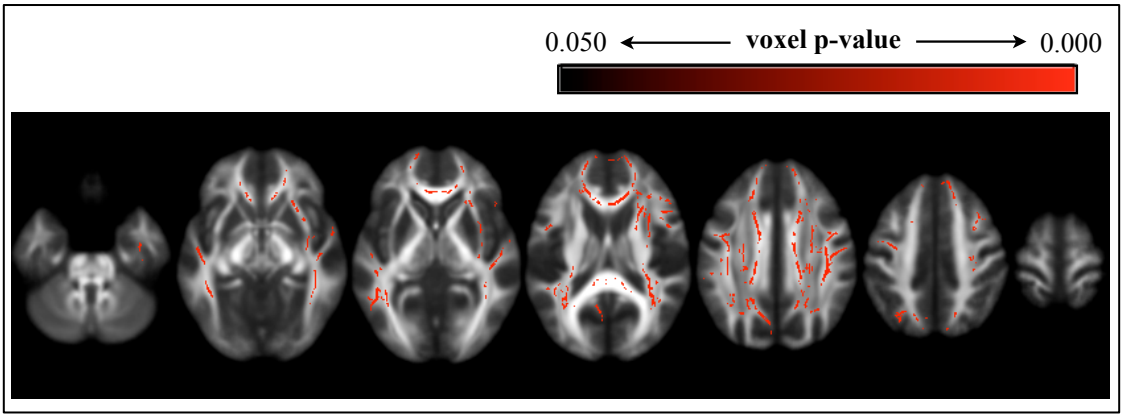
3.3.3 Structural MRI

Using a paired t-test, no significant GM differences, either increases or decreases, were observed for the FEP versus HC subject, ARMS versus HC subject, or FEP versus ARMS subject comparisons respectively (p>0.05, FWE corrected).

3.3.4 Diffusion Tensor Imaging

Using a paired t-test, significant FA reductions were observed in the ARMS group relative to the HC group (see Fig. 3.2). These differences were widespread through the WM skeleton, including the inferior and superior longitudinal, and inferior fronto-occipital fasciculi bilaterally, the right superior fronto-occipital fasciculus and external capsule, the anterior cingulate bilaterally, the right posterior cingulate, the forceps minor bilaterally, the anterior, posterior and superior corona radiata bilaterally, the right anterior thalamic radiation, and the genu, splenium and body of the corpus callosum. In comparison, no significant FA differences were observed between FEP and HC subjects or between FEP and ARMS subjects (p>0.05, FWE corrected).

Figure 3.2. White matter regions with significantly reduced FA values in ARMS versus HCs (p<0.05, FWE corrected; shown in red) (left to right: axial slices with MNI Z-coordinate -28, -6, 2, 16, 32, 46, 67).



3.3.5 Functional MRI

Using a paired t-test analysis no significant differences in functional activation for any of the five functional contrasts (Suppression>Initiation, Initiation>Overt oral repetition of the word 'REST' during Initiation, Suppression>Overt oral repetition of the word 'REST' during Suppression, Initiation>Cross-fixation during Initiation, Suppression>Cross-fixation during Suppression) were observed for FEP versus HC subject, ARMS versus HC subject, or FEP versus ARMS subject comparisons respectively (p>0.05, FWE corrected).

3.3.6 Neuropsychological Profile

Using a paired t-test, both FEP and ARMS subjects performed significantly worse (p<0.05, Bonferroni corrected) than HCs with respect to multiple domains of the CVLT-II including level of recall, level of delayed recall, and number of recall errors. Relative to HCs, FEP subjects also had significantly (p<0.05, Bonferroni corrected) reduced subjective clustering; a learning characteristic, whilst ARMS had significantly worse delayed recognition. A direct comparison showed ARMS subjects performed

marginally worse than FEP subjects in terms of extent and severity, with ARMS subjects performing significantly worse ($p<0.05$, Bonferroni corrected) on a single subcomponent of the learning characteristics domain (see Table 3.3).

Table 3.3**.** Results of paired t-test analyses examining subcomponents of the CVLT-II with respect to each diagnostic comparison.

	CVLT-II Subcomponent	ARMS versus HC	FEP versus HC	FEP versus ARMS
A	**Trial 1 (ss)**	t=-2.999, p=0.008*	t=-2.615, p=0.018*	t=1.262, p=0.228
	Trial 1-5 (ss)	t=-4.106, p=0.001**	t=-3.755, p=0.001**	t=0.264, p=0.795
	Trial B (ss)	t=-2.996, p=0.008*	t=-2.683, p=0.015*	t=-1.700, p=0.111
B	**SD FR (ss)**	t=-4.080, p=0.001**	t=-3.279, p=0.004**	t=0.086, p=0.932
	LD FR (ss)	t=-3.618, p=0.002**	t=-3.899, p=0.001**	t=0.346, p=0.735
	SD CR (ss)	t=-5.133, p=0.00007**	t=-3.571, p=0.002**	t=1.091, p=0.294
	LD CR (ss)	t=-3.904, p=0.001**	t=-3.261, p=0.004**	t=0.608, p=0.553
C	**PI List B vs. Trial 1 (zsd)**	t=1.206, p=0.243	t=0.000, p=1.000	t=-2.520, p=0.024*
	RI SDFR vs. Trial 5 (zsd)	t=-0.456, p=0.654	t=-0.889, p=0.385	t=-0.202, p=0.843
	LDR LDFR vs. Trial 5 (zsd)	t=-0.515, p=0.613	t=-1.587, p=0.130	t=0.191, p=0.851
	LDR LDFR vs. SDFR (zsd)	t=-0.229, p=0.821	t=-0.637, p=0.532	t=0.511, p=0.617
D	**SemClust (ss)**	t=-1.788, p=0.091	t=-0.539, p=0.596	t=3.371, p=0.005**
	SubjClust (ss)	t=-2.435, p=0.026*	t=-3.104, p=0.006**	t=-1.407, p=0.181
E	**TR (ss)**	t=-1.556, p=0.137	t=3.007, p=0.008*	t=2.614, p=0.020*
	TI (ss)	t=1.219, p=0.238	t=0.759, p=0.457	t=-1.208, p=0.247
	TNCI (raw)	t=1.198, p=0.246	t=1.039, p=0.313	t=-1.583, p=0.136
	TS/S (raw) [a]	-	-	-
	TAL (raw)	t=2.357, p=0.030*	t=0.780, p=0.446	t=-0.612, p=0.550
	TRecD (ss)	t=-3.714, p=0.002**	t=-3.152, p=0.006**	t=0.706, p=0.492
F	**TH (ss)**	t=-2.689, p=0.015*	t=-2.295, p=0.034*	t=0.312, p=0.759
	TRD Hits vs. FP (ss)	t=-4.356, p=0.0004**	t=-2.548, p=0.020*	t=1.671, p=0.117
	SRD Hits vs. List B FP (ss)	t=-4.446, p=0.0003**	t=-2.423, p=0.026*	t=1.355, p=0.197
	SemRD List A vs. List B Shared & PFPs (ss)	t=-3.829, p=0.001**	t=-2.143, p=0.046*	t=1.908, p=0.077
	NRD Hits vs. PFPs & UFPs (ss)	t=-3.074, p=0.007*	t=-2.752, p=0.013*	t=0.978, p=0.344
	TRB (ss)	t=0.268, p=0.792	t=1.228, p=0.235	t=0.590, p=0.565
G	**TRD vs. LDFR (%)**	t=1.189, p=0.250	t=0.563, p=0.580	t=0.842, p=0.414
	TRD vs. LDFR (ss)	t=-1.197, p=0.247	t=1.779, p=0.092	t=1.946, p=0.072

** Bonferroni corrected: p-value<0.05. * Uncorrected: p-value<0.05. A-G; CVLT-II test domains (A; Level of Recall, B; Level of Delayed Recall, C; Recall Contrast Measures, D; Learning Characteristics, E; Recall Errors, F; Delay Recognition Trials, G; Recall/Recognition Contrast Measures), SD/F/C/R; Short Delay/Free/Cued/Recall, LD/F/C/R; Long Delay/F/C/R, LDR; Long delay retention, PI; Proactive Interference, RI; Retroactive Interference, SemClus; Semantic Clustering, SubjClust; Subjective Clustering, TR; Total Repetitions, TI; Total Intrusions, TNCI; Total Non-Category Intrusions, TS/S; Total Synonym/Subordinate Intrusions, TAL; Total Across List Intrusions, TRecD; Total Recall Discriminability, TH; Total Hits, FP; False Positives, SRD; Source Recognition Discriminability, SemRD; Semantic RD, NRD; Novel RD, TRB; Total Response Bias, TRD; Total Recognition Discriminability, ss; standard score, zsd; z-score difference. [a] *t* could not be computed because the standard error of the difference is zero.

3.4 Discussion

In the current study, using a univariate approach I aimed to investigate whether significant group level differences could be established using neuroanatomical, neurofunctional, or neuropsychological data, between either FEP and HC subjects, ARMS and HC subjects, and/or FEP and ARMS subjects. Based on these investigations, I hypothesised that at the group level, significant differences would be observable between:

i) FEP and HC subjects with respect to: a) GM, with reductions evident in multiple cortical and subcortical frontal, temporal and parietal regions of FEP subjects relative to HCs (see section 1.2.2.1.2), b) WM, with reductions evident in multiple areas including the inferior, superior and fronto-occipital fasciculi, the posterior thalamic radiations, the coronae radiatae, the internal capsule, and the corpus callosum of FEP subjects relative to HCs (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in parietal and posterior cingulate regions of HCs relative to FEP subjects based on similar studies of ChSz patients (Schneider, et al., 2011), and, d) neuropsychological performance, with FEP subjects performing significantly worse for all subcomponents of the CVLT-II relative to HCs (see section 1.2.3.2).

ii) ARMS and HC subjects with respect to: a) GM, with reductions evident in cortical and subcortical frontal, temporal and parietal regions of the ARMS subjects relative to HCs (see section 1.2.2.1.2), b) WM, with reductions evident in the inferior, superior and fronto-occipital fasciculi, the posterior thalamic radiations, the coronae radiatae, the internal capsule and the corpus

callosum of ARMS subjects relative to HCs (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in the right caudate and anterior cingulate gyri bilaterally of ARMS subjects relative to HCs (see section 1.2.2.3.2), and, d) neuropsychological performance, with ARMS subjects performing significantly worse for all subcomponents of the CVLT-II relative to HCs (see section 1.2.3.2).

iii) FEP and ARMS subjects with respect to: a) GM, with reductions evident in cortical and subcortical frontal, temporal, parietal and cerebellar regions of FEP relative to ARMS subjects (see section 1.2.2.1.2), b) WM, with reductions evident in the inferior, superior and fronto-occipital fasciculi, external and internal capsules, posterior thalamic radiations, coronae radiatae and corpus callosum of FEP relative to ARMS subjects (see section 1.2.2.2.2), c) neural activation elicited by the HSCT, with reductions evident in the anterior cingulate and caudate, in addition to increases in regions of the default mode network, in FEP relative to ARMS subjects, based on previous results of ChSz patients relative to HCs, and ARMS subjects relative to HCs, respectively (see section 1.2.2.3.2), and, d) neuropsychological performance, with FEP subjects performing significantly worse than ARMS subjects for all subcomponents of the CVLT-II based on previous results showing progressive cognitive decline associated with conversion from ARMS to FEP (see section 1.2.3.2)

Contrary to hypotheses 1a-c, no significant group level differences were seen between FEP and HCs with respect to GM, WM integrity or neurofunction. However, CVLT-II scores for FEP subjects were significantly reduced compared to HCs across a range of dimensions consistent with hypothesis 1d. With respect to hypotheses 2a-d, the results were inconsistent with the first and the third of these, with no significant differences seen between ARMS and HC subjects with respect to either GM or neurofunction. However, consistent with hypotheses 2b and 2d standard analysis did detect multiple areas of reduced WM integrity (see Fig. 3.2), and significant CVLT-II performance deficits across a range of dimensions (see Table 3.3), in ARMS subjects relative to HCs. Finally, in disagreement with hypotheses 3a-c, no significant differences were detected between FEP and ARMS subjects with respect to GM, WM integrity or neurofunction. Furthermore, contrary to hypothesis 3d, ARMS subjects in fact performed marginally worse than FEP subjects, however, this was with respect to only a single component of a single test domain (see Table 3.3).

Together, the absence of significant GM or neurofunctional alteration for any of the three diagnostic comparisons, FEP versus HCs, ARMS versus HCs, or FEP versus ARMS is inconsistent with the majority of studies published to date. This is also true with respect to the lack of observed WM alteration between FEP and HC subjects and FEP and ARMS subjects. Outlined in the introduction, and providing the basis for hypotheses 1, 2 and 3, these past studies have suggested FEP and ARMS subjects to be associated with significant alterations in multiple cortical and subcortical brain regions evident in GM, WM and neurofunction (see section 1.3.2). However in accordance with these studies, the results did show that ARMS subjects had multiple areas of significantly reduced WM integrity, relative to HCs, encompassing spatially extensive brain regions which were also consistent with those observed in previous studies

(Carletti, et al., 2012). Furthermore, as hypothesised, FEP and ARMS subjects performed significantly worse than HCs with respect to the CVLT-II and, also, the behavioural component of the HSCT (Fig. 3.1, Table 3.2). This is consistent with previous work that suggests deficits in neuropsychological performance are a characteristic feature of both established psychosis and also its preceding prodrome (Brewer, et al., 2005; Sponheim, et al., 2010; Wood, et al., 2007a). Moreover, the CVLT-II results also showed that the performance of FEP and ARMS subjects were effectively the same, in contrast to their comparison with HCs, supporting the notion that cognitive deficits similar to those evident in FEP may also be apparent in those with an ARMS (Wood, et al., 2007a).

In consideration of those null results highlighted, there are a number of possible interpretations one can make. The first of these is that there were simply no gross differences to detect given the specific FEP, ARMS and HC subjects used in the current study. With respect to neuroanatomy, this is partially consistent with previous studies that found no reliable group level differences in FEP, or ARMS, relative to HC subjects with respect to GM (Velakoulis, et al., 2006; Wood et al., 2005), or in FEP relative to HC subjects with respect to WM (Friedman, et al., 2008; Peters, et al., 2008). Comparatively, the absence of significant neurofunctional differences, in the context of previous studies which have largely reported significant differences in FEP and ARMS subjects relative to HCs (Egerton, et al., 2011), is harder to explain. If we consider past instances where the functional HSCT paradigm has been applied either to ARMS (Allen, et al., 2010) or ChSz (Schneider, et al., 2011) subjects relative to HCs, however, it is noteworthy that there was a discrepancy in the location and direction of the neural correlates reported by each study respectively. Given this discrepancy, it is possible that alterations in the neurofunctional correlates of the HSCT are qualitatively different

between the prodromal and established stages of psychosis, not necessarily proceeding in a manner wholly concordant to that seen in neuroanatomy. One speculative explanation for these contrasting results is that the neurofunctional difference observed by Allen et al. represented a form of neurofunctional compensation in the ARMS subjects allowing them to perform on a par with HCs behaviourally. In contrast, the neurofunctional alterations observed by Schneider et al., which were quite distinct from those observed in ARMS subjects, were associated with significant deficits in task performance and may therefore have reflected aberrant neural processes. Based on this reasoning, it is possible that the ARMS subjects in the current study were relatively more akin to ChSz patients than those ARMS subjects recruited by Allen et al., as reflected by their deficits in task performance and corresponding absence of compensatory neurofunctional alteration. A similar line of reasoning may also explain the absence of neurofunctional differences in this study's FEP subjects supported by the fact that, at least on a psychopathological scale, ARMS and FEP subjects were qualitatively similar on the day of scanning, with neither group suffering symptoms as severe as those typically observed in ChSz (see Table 3.1).

An alternative interpretation of the null-results is that although true effects of group existed, the analysis used here was unable to detect them. One potential explanation for this could be the inherent clinical heterogeneity of the FEP and ARMS cohorts, with each subject having a combination of symptoms unique to themselves – which, though subtly different to other individuals, may warrant the same categorisation (i.e. ARMS or FEP). Such clinical heterogeneity may result in high levels of neuroanatomical and neurofunctional inter-subject variability which in turn would increase standard deviation and decrease statistical power. It may also explain the fact that neuropsychological deficit could apparently occur in the absence of neuroanatomical and neurofunctional

change, since heterogeneous changes in the latter may still give rise to common changes in the former (i.e. common cognitive deficits may emerge from differing alterations in neural processes or neuroanatomy) (Wilkinson & Halligan, 2004).

A third possible explanation for the null-results is that the sample size used in the current study which, relative to recent multi-centre studies for example (Mechelli, et al., 2011) is comparatively small, was not sufficiently powered to detect an effect of interest. However, a recent analysis of effect size in classical inference suggests that, in order to optimize the sensitivity to large effects while minimizing the risk of detecting trivial effects, the optimum sample size for a study is 16-32 (Friston, 2012) (see section 2.1.7.1 for more detail).

In consideration of the HSCT results specifically, it also possible that methodological factors contributed to the lack of neurofunctional results. For example, in the study by Allen et al. where the fMRI version of the HSCT was used to compare ARMS and HC subjects, errors trials were modelled as a separate regressor within the analysis. However this was not done in the present analysis so as to remain consistent with the data used for multivariate analysis, for which classification was the primary focus rather than detection of the neural correlates of subtle cognitive processes *per se*. As such, it is possible that variation during the error responses, equating to noise in the subsequent analysis, reduced the chance of detecting a reliable group level effect. However, the fact that significant behavioural differences were observed in the current study but not by Allen et al. reduces the likelihood that similar, possibly compensatory, neurofunctional changes would have anyway been seen in the event errors were modelled as a separate regressor.

A final possible interpretation of the data supported by the neuropsychological results, is that FEP and ARMS subjects did in fact differ from HCs, and each other, with respect to those data types and comparisons for which no significant differences were observed, but that these differences were characterised by subtle, and diffuse patterns of alteration instead of localised foci of difference. If this were the case, such alterations would be better detected using a multivariate analysis of the kind explored in the next chapter, rather than a standard analysis specifically geared to finding gross group level differences.

3.4.1 Limitations

The fact that the majority of FEP subjects were medicated represents the main limitation of the study. Whilst preliminary evidence suggests antipsychotics do have an impact on both neuroanatomy (Navari & Dazzan, 2009) and neurofunction (Fusar-Poli et al., 2007a), the precise extent and nature of such change remains unclear, and as such, it remains a potential confound with respect to the findings observed here; indeed, it is an issue not specific to the current study, but one that is applicable to the majority of studies of psychiatric patients. Secondly, as a cross sectional study it was not possible to provide a longitudinal analysis identifying group level changes in the ARMS subjects arising over time. Finally, it should be noted that the HSCT is a relatively complex task both to understand, and to perform, and as such, for those subjects with a higher degree of cognitive and functional decline, a less complex task may have been more appropriate.

3.4.2 Conclusion

To conclude, inconsistent with the majority of literature, the findings here suggest that ARMS and FEP subjects do not significantly differ from HCs, or each other, in a manner detectable by standard analysis, either in terms of GM or neurofunction, nor FEP from ARMS or HC subjects in terms of WM. Consistent with previous studies however, standard univariate analysis was able to detect widespread reduction in WM integrity in ARMS relative to HCs, and also significant neuropsychological deficits in FEP and ARMS subjects relative to HCs, as well as FEP relative to ARMS subjects. Given the presence of impaired neuropsychological performance in ARMS and FEP relative to HC subjects, it is possible corresponding changes in neuroanatomy are in fact present where none were detected. Such changes may, however, be characterised by subtle and diffuse patterns of alteration as opposed to localised gross differences. In this case, it is possible a multivariate analysis would be more sensitive to such alterations, the results of which are explored in the next chapter.

Chapter 4

Using Genetic, Cognitive and Multi-Modal Neuroimaging data to identify the At-Risk Mental State and First Episode Psychosis at Individual level

4.1 Introduction

Considerable effort has been made over the last 30 years to identify biological and cognitive markers of schizophrenia. A large number of studies report significant differences in chronic schizophrenia (ChSz) patients relative to healthy controls (HCs) across a range of neurobiological and neurocognitive measures, including structural, functional and diffusion tensor MRI (sMRI, fMRI, DTI) (Ellison-Wright & Bullmore, 2009; Ellison-Wright et al., 2008; Minzenberg, et al., 2009; Pettersson-Yeo, et al., 2011), genotype (Ripke, et al., 2011; Steinberg, et al., 2011) and neuropsychological profile (Minzenberg, et al., 2009; Tyson et al., 2004) (see section 1.2). More recently, efforts to facilitate earlier and more effective treatment intervention have resulted in studies focusing on those in the earliest stages of the illness, namely, individuals with a first-episode of psychosis (FEP) and those assessed as having an at-risk mental state (ARMS). In these groups, similar neuroanatomical, neurofunctional and cognitive alterations (Allen et al., 2011; Allen, et al., 2010; Benetti, et al., 2009; Bilder, et al., 2000; Crossley, et al., 2009; Fusar-Poli, et al., 2012b; Keefe et al., 2006; Koutsouleris et

al., 2010; Mechelli, et al., 2011; Seidman et al., 2010; Walterfang et al., 2008; Wood, et al., 2007a) that may, at least in part, be mediated genetically (Fusar-Poli, et al., 2012b), have also been reported, though such alterations are usually less severe than those seen in ChSz groups (Egerton, et al., 2011) (see section 1.3). The majority of such studies have, however, largely employed univariate analyses that allow inference at the group level only. To promote the clinical translation of such work therefore efforts have progressively turned toward alternative analytical approaches that allow inference at the level of the individual.

One such technique is the supervised learning method Support Vector Machine (SVM). A type of multivariate pattern recognition algorithm, SVM has become increasingly used in studies of psychiatric and neurological disorder (Orrù, et al., 2012), the rationale for which is twofold: firstly, SVM allows inference at the single-subject, rather than group, level (Norman et al., 2006); secondly, as a multivariate analysis, SVM is able to account for the inter-relationship between different within-modality measures (i.e. features) for each subject by considering them simultaneously (Lao, et al., 2004). Specifically, SVM involves the development of a generalised decision function (represented by an optimal separating hyperplane (OSH)) using a known "training" data set (e.g. voxel intensities), able to discriminate between examples (i.e. subjects) belonging to two predefined classes (e.g. diagnostic categories). This function is then applied to new, as yet unseen "test" data, and it's accuracy assessed in terms of the proportion of examples correctly classified providing an estimate of how well the classifier can be expected to generalise to future individual cases (Burges, 1998; Schölkopf & Smola, 2002) (see section 1.3 for more detail).

4.1.1 Previous SVM Studies of Psychosis using Uni-modal Data

At present, only a handful of studies have used SVM to investigate psychosis, with those that have predominantly employing data from only a single modality. Studies using sMRI for example, report that ChSz patients can be significantly discriminated from HCs with accuracies of 81.1% (Davatzikos, et al., 2005), FEP from HC subjects with 86.1% (Sun, et al., 2009), and, ARMS subjects from HCs with an accuracy of 82% (Koutsouleris, et al., 2009a). DTI has also been used by one study which reported that ChSz patients can be discriminated from HCs with 90.62% accuracy based on white matter integrity (Ingalhalikar, et al., 2010). In addition two studies using fMRI data reported the ability to discriminate ChSz patients from HCs with accuracies of 92.4% (Costafreda, et al., 2011) and 81.6% (Yang, et al., 2010) respectively. By contrast, only one study has applied SVM to genotype data in the context of psychosis, and reported that genetic information could accurately discriminate ChSz subjects from HCs with 73.9% accuracy (Yang, et al., 2010). Lastly, neuropsychological profile has also been employed by one recent study which reported that ARMS subjects were discriminable from HCs with 94.2% accuracy (Koutsouleris, et al., 2011b). Taken together, these studies support the notion that the use of individual modalities may allow discrimination at the single-subject level between those with established psychosis and HC subjects, and between those at clinically high-risk and HC subjects. However, as no investigation has yet gathered data from such a wide range of modalities during the same study, the relative accuracies of genetic, sMRI, DTI, fMRI and neuropsychological data within the same population(s) is unknown. Furthermore, since the majority of work so far has applied predominantly to those with ChSz, it is less clear whether these same metrics can be reliably utilised to draw inference at individual level for FEP and ARMS

subjects, differentiating them either from HCs, or, for the first time at cross-sectional level, from each other.

4.1.2 Aims and Hypotheses

In the current study, I aimed to investigate the discriminative potential of genetic, sMRI, DTI, fMRI and/or neuropsychological data in the classification of FEP, ARMS and HC subjects at individual level. Based on the evidence currently available, my hypotheses were as follows;

i) FEP subjects would be discriminable from HCs using the most types of data whilst in comparison, fewer data types would be able to discriminate ARMS from HC subjects. Specifically, I hypothesised that fMRI and genotype would be those data types least likely to successfully differentiate ARMS subjects from HCs, given that, a) with respect to neural activation, group level results for ARMS subjects have proved relatively less consistent with mixed reports of both increases and decreases relative to HCs, in addition to the magnitude of such differences typically less than the difference between FEP and HC subjects (Benetti, et al., 2009; Crossley, et al., 2009; Egerton, et al., 2011), and b) containing a proportion of individuals who will never transition to psychosis and who may also have no inherent familial risk, it is logically less likely that a common pattern of SNPs associated specifically with established psychosis will be found that is shared by all ARMS subjects.

ii) FEP and ARMS subjects would be directly discriminable with structural MRI and neuropsychological data being the most able to differentiate the two, based on the relatively greater consistency and robustness of the group

level findings of these two modalities when comparing FEP and ARMS subjects to HCs, in comparison to findings from fMRI and genetics studies (see section 1.2).

4.2 Materials and Methods

A brief description of the materials and methods is provided in this section. For further methodological details, please see sections 2.1.5 and 3.2.

4.2.1 Participants

4.2.1.1 First Episode Psychosis

Nineteen subjects were recruited through the South London and Maudsley National Health Service Trust (http://www.slam.nhs.uk). All had experienced a FEP within the past 24 months that met DSM-IV-TR criteria for a schizophreniform psychosis.

4.2.1.2 At-Risk Mental State

Nineteen subjects were recruited from OASIS (Outreach and Support in Southeast London), a clinical service for young people at high-risk of developing psychosis (Broome, et al., 2005). Their clinical status was defined by a trained clinical psychiatrist according to the PACE (Personal Assessment and Crisis Evaluation) criteria (Yung, et al., 1998), and their status confirmed using the CAARMS (Comprehensive Assessment of At-Risk Mental States) (Phillips, et al., 2000).

4.2.1.3 Healthy Controls

Twenty-three subjects were recruited from the local area through advertising. No subjects met criteria for a DSM-IV-TR psychiatric disorder, fulfilled the PACE criteria for prodromal symptoms nor had a first-degree family history of psychiatric disorder (see sections 2.1.5 and 2.2.3 for more detail).

All subjects included in the study were 18-35 years old and spoke English as their first language. Exclusion criteria included a history of neurological disorder, evidence of DSM-IV-TR criteria for substance abuse or dependence, significant visual or hearing impairment, prior head trauma resulting in loss of consciousness and/or hospitalisation, or any contraindication to magnetic field exposure (e.g. pregnancy, pacemaker, etc.).

4.2.1.4 Subject Pairing

For reasons discussed in sections 2.1.6 and 2.5.3 for the purpose of SVM implemented through PROBID (http://www.brainmap.co.uk/probid.htm), balanced numbers of subject pairs, matched here for age and gender, are required for each diagnostic comparison (e.g. ARMS versus HCs). Using these pairing criteria, from our cohort of 19 FEP, 19 ARMS and 23 HCs, 19, 19 and 15 subject pairs were available for the FEP versus HC, ARMS versus HC, and FEP versus diagnostic comparisons respectively, matched for age (±4 years) and gender. For demographic details of each diagnostic comparison please see Table 3.1 in section 3.2.

4.2.2 Data Acquisition

4.2.2.1 Magnetic Resonance Imaging

All neuroimaging was conducted using a 3T MRI scanner (Signa, LX-GE, Milwaukee, USA) at the Maudsley Hospital, London. For the current study, sMRI, DTI and fMRI data was acquired for each subject during the same scanning session. For a full description of the acquisition sequence and parameters used for each scan, please refer to section 3.2.2.3.

4.2.2.2 Neuropsychological Profile

The second edition of California verbal learning test (CVLT-II) was administered to each subject by a trained researcher during the pre-scan assessment interview (see section 2.2) and the subject's scores recorded.

4.2.2.3 Molecular Genetics

Saliva samples were obtained from each subject with informed consent, using the Oragene® DNA collection kit (DNA Genotek Inc, Ontario, Canada), preceded by half an hour of nil by mouth. DNA was manually extracted as per the recommended protocol established by DNA Genotek Incorporated (www.dnagenotek.com/ROW/pdf/PD-PR-006.pdf).

4.2.3 Data Analysis

4.2.3.1 Structural MRI

Structural images were preprocessed using the Diffeomorphic Anatomical Registration using the Exponentiated Lie algebra (DARTEL) toolbox (Ashburner, 2007) in SPM8 software (http://www.fil.ion.ucl.ac.uk/spm) running under Matlab7.1 (Math Works, Natick, MA, USA). This procedure involves the creation of a study-specific template and the segmentation of each individual image using said template, with the aim of maximizing accuracy and sensitivity (Yassa & Stark, 2009) (see section 3.2.3.3 for more detail).

4.2.3.2 Diffusion Tensor Imaging

The diffusion data was preprocessed using the ExploreDTI (Leemans, et al., 2009) software package, including the RESTORE (Chang, et al., 2005) algorithm, in order to generate FA maps corrected for eddy current distortion, head motion, b-matrix reorientation and rejection of data outliers. These images were then used to create FA 'skeletons' depicting each subject's unique WM network and associated FA value defined integrity for each voxel, using Tract-Based-Spatial-Statistics (TBSS) software (Smith, et al., 2006) (see section 3.2.3.4 for more detail).

4.2.3.3 Functional MRI

Functional images were pre-processed using SPM8 software (http://www.fil.ion.ucl.ac.uk/spm) running under Matlab7.1 (Math Works, Natick, MA, USA). Following the standard SPM8 functional imaging pipeline for preprocessing and analysis, using the parameter estimates obtained for all brain voxels from the task's six experimental conditions (1) initiation; 2) suppression; 3/4) repetition of 'REST' during initiation/suppression; and 5/6) cross-fixation during initiation/suppression, five contrasts of interest were computed, namely; Suppression>Initiation, Initiation>Repetition of 'REST' during Initiation, Suppression>Repetition of 'REST' during Suppression, Initiation>Cross-Fixation during Initiation, Suppression>Cross-Fixation during Suppression (see sections 2.4.2.2 and 3.2.2.2.3 for detail). Since the primary objective was to identify the potential for patient classification based on functional activity and not to make inference regarding the mechanisms underlying language processing per se, error responses were not removed from the initiation or suppression conditions nor modeled as a separate regressor.

4.2.3.4 Neuropsychological Profile

Each subjects recorded scores were inputted into the associated CVLT-II software package (Delis & Fridlund, 2000) that provides a comprehensive summary of raw and demographically corrected standardised scores for the different task components measured by the test. This resulted in a total of 45 separate scores, representing 7 different test domains (see section 2.2.2.2, and Table 4.1 for more detail).

4.2.3.5 Molecular Genetics

DNA extracted from each subject was genotyped by KBioscience (Herts, UK) (www.kbioscience.co.uk) for a pre-selected list of 26 psychosis associated SNPs (see Table 1 in section 2.3), using the KASP genotyping system. These 26 represent a complete list of all SNPs showing a positive genome-wide association with schizophrenia and/or bipolar disorder, published (or *in press*, to the best of my knowledge) until June 14th 2011. Observed and Hardy Weinberg equilibrium-expected frequencies were compared for each SNP using a Fisher's exact test implemented through the web-based statistical software, http://vassarstats.net/fisher2x3.html. As per standard practise (Lunetta, 2008), this was calculated in the HC group in the most prevalent ethnicity, British Caucasian. All SNPs were compliant with HWE (Fisher's exact test: $p>0.05$).

Table 4.1. CVLT-II output scores for each subject used as input data for SVM

Test Domain	**CVLT-II Subcomponent**
Level of Recall	1) trial 1 rs 2) trial 1 ss 3) trial 1-5 rs 4) trial 1-5 ss 5) trial B rs 6) trial B ss
Level of Delayed Recall	7) short delayed free recall rs 8) short delayed free recall ss 9) long delayed free recall rs 10) long delayed free recall ss 11) short delayed cued recall rs 12) short delayed cued recall ss 13) long delayed cued recall rs 14) long delayed cued recall ss
Recall Contrast Measures	15) proactive interference z-score difference zsd 16) retroactive interference zsd 17) long delayed free recall versus trial 5 zsd 18) long delayed free recall versus short delayed free recall zsd
Learning Characteristics	19) semantic clustering rs 20) semantic clustering ss 21) subjective clustering rs 22) subjective clustering ss
Recall Errors	23) total repetitions rs 24) total repetitions ss 25) total intrusions rs 26) total intrusions ss 27) total non-category intrusions rs 28) total synonym/subordinate intrusions rs 29) total across list intrusions rs 30) total recall discriminability rs 31) total recall discriminability ss
Delayed Recognition Trials	32) total hits rs 33) total hits ss 34) total recognition discriminability rs 35) total recognition discriminability ss 36) source recognition discriminability rs 37) source recognition discriminability ss 38) semantic recognition rs 39) semantic recognition ss 40) novel recognition discriminability rs 41) novel recognition discriminability ss 42) total response bias rs 43) total response bias ss
Recall/Recognition Contrast Measures	44) total recognition discriminability percentage score 45) total recognition discriminability ss

rs: raw score, ss: standardized score, zsd: z-score difference.

4.2.3.6 Support Vector Machine

As described in section 2.5, SVM is a supervised machine learning algorithm, that seeks to learn a decision function that correctly predicts the class label for each data point, based on a set of training examples, that can then be used to predict the labels for a set of unseen testing examples (Burges, 1998). To achieve this, the processing pipeline for SVM involves three main steps; feature extraction, classifier training and testing, and classifier evaluation.

4.2.3.6.1 Feature Extraction

Feature extraction involves converting data from its original form into a set of 'features' that can be entered into the SVM. This requires each subject's input data to be represented as a single column vector of values, where each element corresponds to one feature (e.g. for sMRI input data, each feature is the signal intensity for a given voxel).

Using the PROBID (www.brainmap.co.uk/probid.htm) software package, feature extraction for neuroimaging data, as used here (sMRI, DTI and fMRI), is performed automatically. In this case, the input data entered for each subject into the SVM for the three modalities respectively were, the sMRI T1-weighted segmented GM image, the DTI fractional anisotropy (FA) skeleton image, and each individual fMRI contrast.

For non-neuroimaging data in comparison, the feature extraction process must be performed manually. For the CVLT-II data this involved creating a column vector of the total range of output scores generated by the associated software package (Delis & Fridlund, 2000) for each subject. Both raw and standardised scores were included given that the primary motivation was to investigate whether CVLT-II data can be used to classify individuals based on a pattern of alterations, rather than group level differences

in specific test subcomponents/domains. Critically, these scores were entered in the same order for each subject (outlined in Table 4.1), such that each column element corresponded to the same CVLT-II subcomponent score for every subject. These column vectors were then used as input data for the SVM.

With respect to genetic data, the feature extraction process was the same as used for the CVLT-II data, with one additional step. Specifically, the genotype of each SNP for each subject was first orthogonally coded – e.g. genotypes AA, AB, BB would be coded '1 0 0', '0 1 0' and '0 0 1' respectively – before the values for each subject were collated into a single column vector with the SNPs in the same order for each subject. This orthogonal coding scheme was employed to prevent introducing an artificial ordinal relationship between genotypes. In cases where one or more SNPs could not be genotyped for a given subject, these were excluded for all other pairs in the SVM comparison since each vector length must be the same (see section 2.5.1). The number of SNPs therefore entered into the SVM for FEP versus HC, ARMS versus HC and FEP versus ARMS were 20, 20 and 19 respectively.

4.2.3.6.2 Classifier Training

Each subject's data (segmented GM images, FA skeletons, HSCT contrast image, orthogonally coded genotype data or CVLT score vectors) were separately entered into SVMs (Burges, 1998) as implemented in the PROBID software package (http://www.brainmap.co.uk/probid.htm) running under Matlab7.1 (Math Works, Natick, MA, USA) in order to assess the diagnostic potential of each modality with respect to ARMS and FEP subjects relative to HCs, and also to each other. For each comparison subject pairs matched for age (±4 years) and gender were used to construct samples for the classifier, with each individual scan, genotype or neuropsychological score input

vector, treated as a data point located in high dimensional space and assigned by the operator to a given class (e.g. FEP or HC). SVM comparator groups comprised 19, 19 and 15 subject pairs for FEP versus HC, ARMS versus HC, and FEP versus ARMS respectively. Since one ARMS subject did not complete the CVLT-II however, only 18 SVM subject pairs were examined for the ARMS versus HC CVLT-II based comparison. Furthermore, due to 4 FEP, 1 ARMS and 1 HC subject(s) who declined to provide a saliva sample, only 14, 17 and 12 subject pairs were used as SVM input for the FEP versus HC, ARMS versus HC and FEP versus ARMS genotype comparisons respectively.

Each classifier was embedded in a leave-one-out cross-validation (LOOCV) framework, whereby all input vectors except those from one pair (one subject from each group) were used as training data for the classifier and the remaining pair withheld as test data (see section 2.5.2 for more detail).

4.2.3.6.3 Classifier Performance Evaluation

The accuracy of the classifier was calculated by taking the mean of its sensitivity and specificity (Hastie, 2001) (see Box 1) across all LOOCV folds; where true positives and negatives denote the number of class 1 (e.g. FEP), and class 2 (e.g. HC), subjects correctly classified as class 1, and class 2, respectively, whilst false positives and negatives denote the number of class 1, and class 2, subjects incorrectly classified as class 2, and class 1, respectively. Statistical significance of the accuracy was determined by a permutation test, whereby subjects were randomly assigned to a class and the LOOCV cycle repeated 1000 times. This provided a distribution of accuracies reflecting the null hypothesis that the classifier did not exceed chance. The number of times where

it was greater than or equal to the true accuracy was then divided by 1000 to estimate a p-value for the accuracy.

Box 1.

$$\textbf{Sensitivity} = \frac{\text{Number of True Positives}}{\text{Number of True Positives} + \text{Number of False Negatives}}$$

$$\textbf{Specificity} = \frac{\text{Number of True Negatives}}{\text{Number of True Negatives} + \text{Number of False Positives}}$$

$$\textbf{Accuracy} = \frac{\text{Number of True Positives} + \text{Number of True Negatives}}{\text{All Subjects}}$$

For each neuroimaging comparison a discrimination map was produced visualising each voxel's weight-vector score (w_i) – representing its relative contribution in defining the OSH – displaying the pattern of regions able to discriminate each group. For successful genetic and CVLT-II based classifiers, analogous graphs showing the w_i for each SNP or CVLT-II subcomponent respectively were also produced. Unlike previous studies (Marquand, et al., 2008; Mourão-Miranda, et al., 2005) no map threshold was applied since any successful OSH was founded on the total number of voxel intensities, SNPs, or CVLT-II scores entered into the SVM. Moreover, since a feature selection step was not employed, nor *a priori* regions of interest specified, it was also not be possible to draw inferences regarding specific regions, SNPs, nor CVLT-II subcomponents out of the context of the overall pattern, in contrast to mass-univariate results (Brammer, 2009). This is due to the fact that the weight vector score assigned to each feature is only done so in the context of every other feature in the pattern (see section 2.5.4 for further detail).

For all classifiers, a linear kernel was used, allowing direct extraction of the weight vector, and the SVM parameter ***C*** (which regulates the balance between maximising the margin between data points and allowing misclassification) was fixed at one for all cases (default value) (for a detailed description of the general framework of SVM please refer to Schölkopf & Smola (2002) and/or Burges (1998)).

4.2.3.6.4 SVM Classification Accuracies

In order to correct for multiple comparisons both a Holm-Bonferroni step-down procedure, which controls for family wise error (FWE) (Holm, 1979), in addition to the generally less conservative Benjamini-Hochberg procedure, which controls for false discovery rate (FDR) (Benjamini & Hochberg, 1995) were employed. However, since both procedures are intended for independent data, which the comparisons here are unlikely to be, there was an increased risk of type II error. Therefore, in the absence of an optimally established method for correcting non-independent hypotheses, for completeness both corrected and uncorrected accuracies are reported (see Table 4.2).

4.2.3.6.5 Comparing Classifiers

Though the study's primary focus was to investigate whether each data type can, or cannot, successfully classify FEP and ARMS subjects from HCs, and/or each other, for completeness, I also performed a non-parametric Cochran's Q test to examine whether the levels of accuracy for each classifier differed significantly for each diagnostic comparison. In the event of a significant result, a post-hoc McNemar's test, with Bonferroni correction, was then applied to identify which specific classifiers were statistically different.

4.3 Results

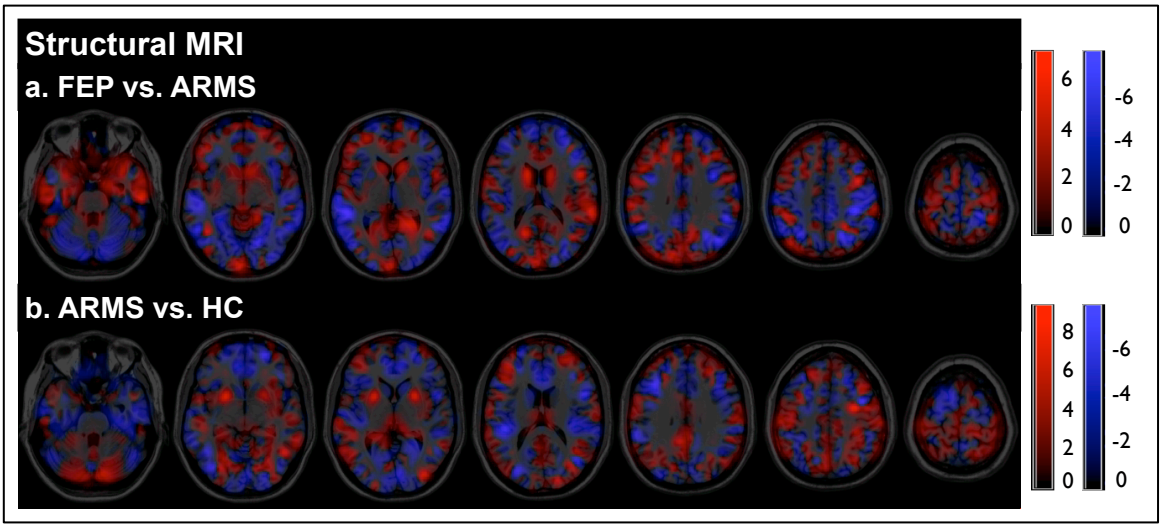
4.3.1 Demographics

As reported in chapter 3, there were no significant differences with respect to premorbid-IQ between any of the groups (p>0.05), nor was there a significant difference in PANSS scores (total, positive, negative or general) between ARMS and FEP subjects (p>0.05). With respect to medication, all FEP subjects, except one, were medicated. In comparison all ARMS subjects were medication-naïve, apart from two (see Table 3.1 in section 3.2 for more detail).

4.3.2 SVM classification of grey matter images

Using GM images, SVM was able to successfully discriminate FEP from ARMS subjects and ARMS from HC subjects with accuracies of 76.67% (p<0.05, FWE corrected) and 68.42% (p<0.05) respectively. At a trend level only, SVM was also able to discriminate FEP from HC subjects with an accuracy of 63.16% (p=0.066) (see Table 2). For the FEP versus ARMS subject comparison the regional pattern most representative of FEP subjects was more rostrally and subcortically concentrated in comparison to the ARMS group. Similarly for the ARMS versus HC comparison, the regional pattern that most typified the ARM group appeared concentrated in more extreme cortical, rostral and caudal regions (Fig. 4.1a-b).

Figure 4.1. Multivariate weight maps for successful classifiers using sMRI data

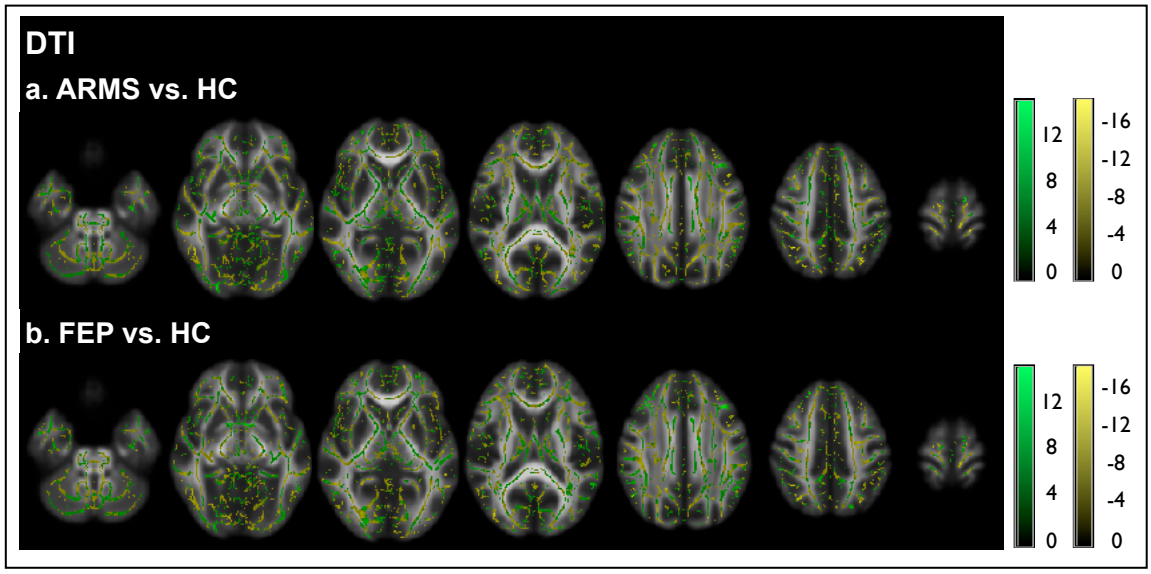


Figures 4.1a-b: Multivariate maps showing the pattern of grey matter regions used to discriminate; (a) FEP and ARMS subjects: Red indicates discrimination in favour of the FEP versus the ARMS group, whilst blue indicates discrimination in favour of the ARMS group versus the FEP group; (b) ARMS vs. HC: Red indicates discrimination in favour of the ARMS versus the HC group, whilst blue indicates discrimination in favour of the HC group versus the ARMS group. ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Control. Figures 4.1a-b: left to right, axial slices with MNI Z-coordinate -28, -6, 2, 16, 32, 46, 67. Colour scales (right) refer to the weight vector score for each voxel representing its relative contribution the optimal separating hyperplane in relation to every other voxel.

4.3.3 SVM classification of FA skeletons

Based on FA skeletons, SVM was able to successfully discriminate both FEP from HC subjects, and ARMS from HC subjects with 65.79% accuracy (p<0.05). The pattern of regions used for each classification was widely and diffusely spread with no clear concentration of regions discernible (Fig. 4.2a-b). In contrast it was not possible to directly discriminate FEP from ARMS subjects using DTI derived FA skeletons with significant accuracy (see Table 4.2).

Figure 4.2. Multivariate weight maps for successful classifiers using DTI data



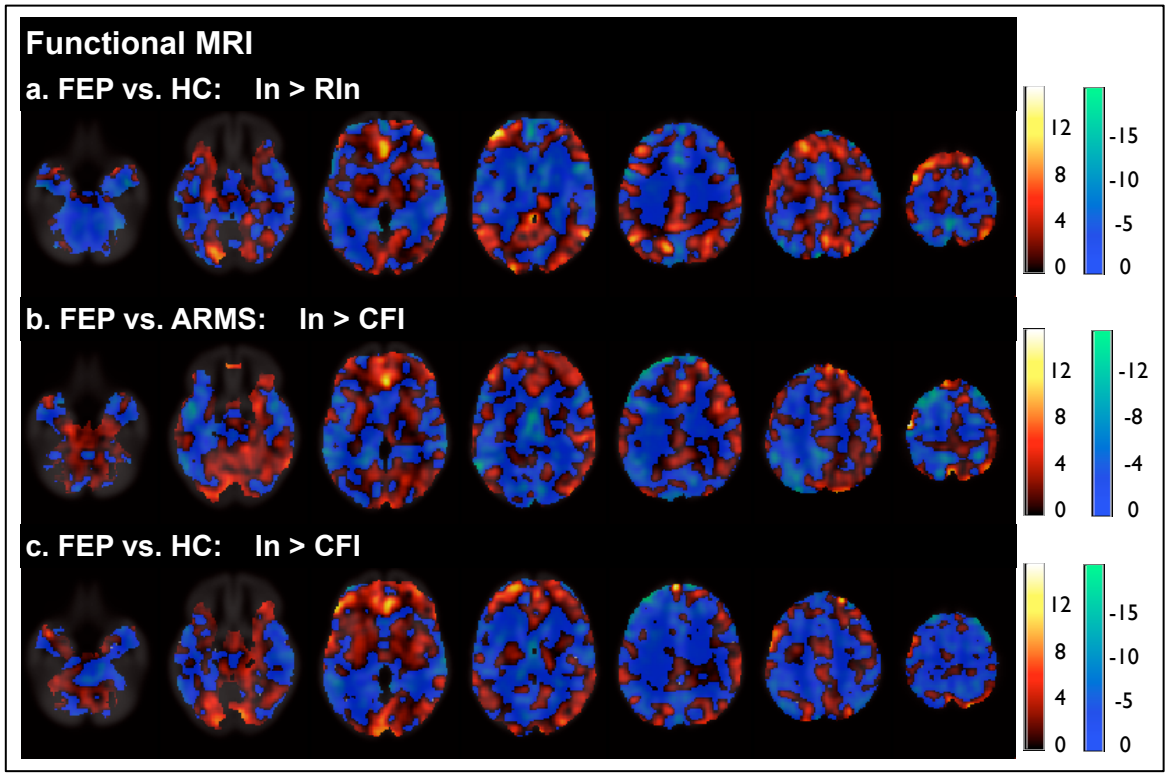
Figures 4.2a-b: Multivariate maps showing the pattern of white matter regions used to discriminate; (a) ARMS and HC subjects: Green indicates discrimination in favour of the ARMS versus the HC group, whilst yellow indicates discrimination in favour of the HC group versus the ARMS group; (b) FEP and HC subjects: Green indicates discrimination in favour of the FEP versus the HC group, whilst yellow indicates discrimination in favour of the HC group versus the ARMS group. ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Control. Figures 4.2a-b: left to right, axial slices with MNI Z-coordinate -28, -6, 2, 16, 32, 46, 67. Colour scales (right) refer to the weight vector score for each voxel representing its relative contribution the optimal separating hyperplane in relation to every other voxel.

4.3.4 SVM classification of HSCT contrasts

Of the five contrasts tested, only two were able to make successful discriminations. Using the Initiation>Repetition of 'REST' during Initiation contrast FEP were distinguishable from HC subjects with an accuracy of 65.79% (p<0.05). Second, using the contrast Initiation>Cross-Fixation during Initiation, SVM could discriminate between both FEP and ARMS, and also between FEP and HC subjects, with accuracies of 73.33% (p<0.05, FDR corrected) and 68.42% (p<0.05) respectively. As figures 4.3a-c show, the regional pattern discriminating FEP from ARMS and HC subjects was concentrated in the frontal and occipital cortices; in contrast the pattern that most

typified the ARMS and HC relative to FEP subjects was widespread with greater prominence in the areas encompassing the central fissure. fMRI data was unable to distinguish ARMS from HC subjects (see Table 4.2).

Figure 4.3 Multivariate weight maps for successful classifiers using fMRI data

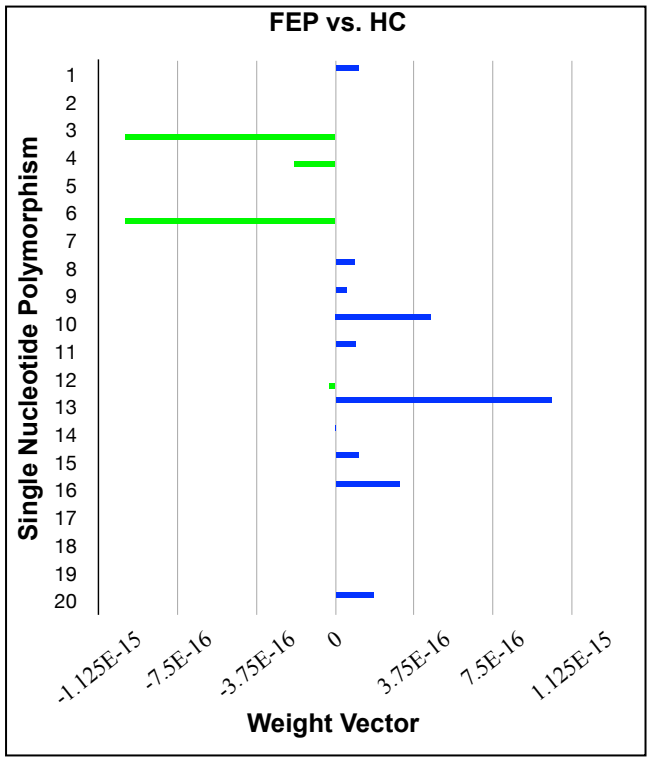


Figures 4.3a-c: Multivariate maps showing the pattern of neurofunction used to discriminate; (a) FEP and HC subjects using the In>RIn HSCT contrast: Gold indicates discrimination in favour of the FEP versus the HC group, whilst turquoise indicates discrimination in favour of the HC group versus the FEP group; (b) FEP and ARMS subjects using the In>CFI HSCT contrast: Gold indicates discrimination in favour of the FEP versus the ARMS group, whilst turquoise indicates discrimination in favour of the ARMS group versus the FEP group; (c) FEP and HC subjects using the In>CFI HSCT contrast: Gold indicates discrimination in favour of the FEP versus the HC group, whilst turquoise indicates discrimination in favour of the HC group versus the FEP group. ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Control, In: Initiation, RIn: Repetition of 'REST' during initiation, CFI: Cross-fixation during Initiation. Figures 4.3a-c: left to right, axial slices with MNI Z-coordinate -28, -6, 2, 16, 32, 46, 67. Colour scales (right) refer to the weight vector score for each voxel representing its relative contribution the optimal separating hyperplane in relation to every other voxel.

4.3.5 SVM classification using Genotype

Using genetic information comprising data from a combination of twenty psychosis associated SNPs, SVM was able to successfully discriminate FEP subjects from HCs with an accuracy of 67.86% (p<0.05) (Fig. 4.4). Comparatively it was not possible to discriminate ARMS from HCs, nor FEP from ARMS subjects using genetic data (see Table 4.2).

Figure 4.4 Weight vector scores for successful Genetic based classifier

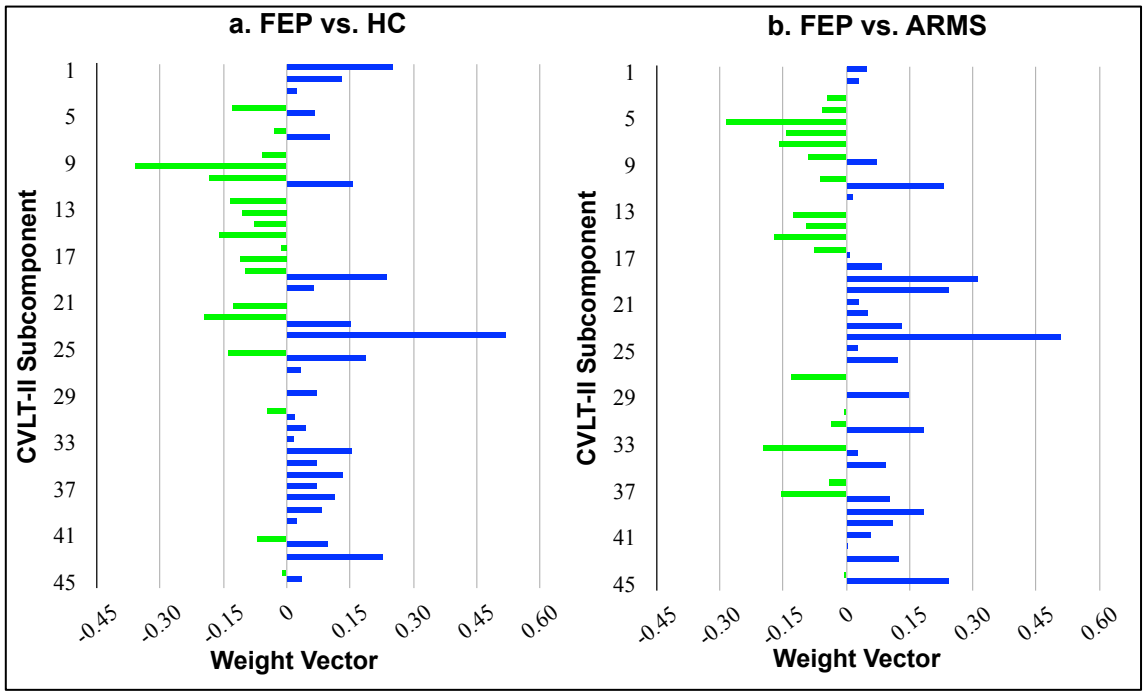


Figures 4.4: Bar charts showing the weight vectors for each SNP, representing their relative contribution to the OSH, used to discriminate FEP and HC subjects: Blue indicates discrimination in favour of the FEP versus the HC group, whilst green indicates discrimination in favour of the HC group versus the FEP group ('E-n': Multiplies the preceding value by $(10)^{-n}$, where n is a real number) ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Control. (SNPs 1-20 correspond to SNPs 1-20 in Table 2.1 in section 2.3).

4.3.6 SVM classification using Neuropsychological Profile

Based on the collated score representing performance over the whole CVLT-II, SVM was able to successfully discriminate both FEP from HC subjects and also FEP from ARMS subjects with accuracies of 73.69% (p<0.05, FDR corrected) and 66.67% (p<0.05) respectively (Figures 4.5a-b). In contrast, CVLT-II score could not accurately differentiate ARMS from HC subjects (see Table 4.2).

Figure 4.5 Weight vector scores for successful CVLT-II based classifiers



Figures 4.5a-b: Bar charts showing the weight vectors for each CVLT-II sub-component, representing their relative contribution to the OSH, used to discriminate; (a) FEP and HC subjects: blue indicates discrimination in favour of the FEP versus the HC group, whilst green indicates discrimination in favour of the HC group versus the FEP group, and (b) FEP and ARMS subjects: blue indicates discrimination in favour of the FEP versus the ARMS group, whilst green indicates discrimination in favour of the ARMS group versus the FEP group. ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Control. (CVLT-II subcomponents 1-45 are shown in Table 4.1).

Table 4.2. Summary of classification accuracies (%) and (p-value) for each binary group comparison, using sMRI, fMRI, genetic or neuropsychological input data in conjunction with SVM.

SVM Comparison	SVM Input Data								
	GM	**FA Skeleton**	**Su > In**	**In > RIn**	**Su > RSu**	**In > CFI**	**Su > CFS**	**Genotype**	**CVLT-II**
ARMS vs. HC (%)	68.42* (68.42/68.42) (p=0.01)	65.79* (68.42/63.16) (p=0.032)	47.37 (31.58/63.16) (p=0.707)	57.89 (57.89/57.89) (p=0.179)	60.53 (57.89/63.16) (p=0.113)	36.84 (36.84/36.84) (p=0.967)	60.53 (57.89/63.16) (p=0.127)	52.94 (52.94/52.94) (p=0.457)	50.00 (44.44/55.56) (p=0.590)
FEP vs. HC (%)	63.16 (57.89/68.42) (p=0.066)	65.79* (68.42/63.16) (p=0.031)	63.16 (47.37/78.95) (p=0.064)	65.79* (63.16/68.42) (p=0.034)	44.74 (36.84/52.63) (p=0.791)	68.42* (63.16/73.68) (p=0.017)	47.37 (42.11/52.63) (p=0.694)	67.86* (71.43/64.29) (p=0.031)	73.69** (68.42/78.95) (p=0.002)
FEP vs. ARMS (%)	76.67*** (80.00/73.33) (p=0.001)	56.67 (46.67/66.67) (p=0.281)	53.33 (40.00/66.67) (p=0.404)	63.33 (53.33/73.33) (p=0.087)	46.67 (46.67/46.67) (p=0.731)	73.33** (66.67/80.00) (p=0.005)	53.33 (40.00/66.67) (p=0.425)	33.33 (41.67/25.00) (p=0.927)	66.67* (66.67/66.67) (p=0.034)

Starred results are significant at ***p<0.05 FWE corrected, **p<0.05 FDR corrected, *p<0.05 uncorrected. GM; Grey Matter, FA Skeleton; Fractional Anisotropy Skeleton, CVLT-II; California Verbal Learning Test – second edition. HSCT contrast conditions - In; Initiation, Su; Suppression, RIn; Repetition of Initiation, RSu; Repetition of Suppression, CFI; Cross-fixation during Initiation, CFS; Cross-fixation during Suppression, FEP: First Episode Psychosis, ARMS: At-Risk Mental State, HC: Healthy Control, SVM: Support Vector Machine.

4.3.7 Comparison of Classifiers

Using a Cochran's Q test no significant differences were observed in the levels of accuracy between the classifiers intended to discriminate ARMS and HC subjects (Q=8.856, p=0.451), nor FEP and HC subjects (Q=10.400, p=0.319). Whilst a significant difference was observed between the classifiers intended to discriminate FEP and ARMS subjects (Q=18.353, p=0.031), subsequent post-hoc McNemar's tests (Bonferroni corrected) comparing individual classifiers were non-significant (p>0.05).

4.3.8 Potential Confounds

Since most FEP subjects were medicated I examined whether any successful classifier able to discriminate them from ARMS, or HC, subjects may have possibly been driven by this potential confound. This was achieved by performing a Pearson's correlation analysis between the projection of each FEP subject's input data onto the weight vector (i.e. the distance of each test subject's scan from the hyperplane, quantifying the relative ease, or difficulty, with which they were categorized) and their corresponding medication measure, i.e. total dose and mean dose per day. No significant correlations were found either for total dose/mean dose per day and i) successful FEP versus HC classifiers based either on genotype (r=0.070, n=14, p=0.811/r=-0.247, n=14, p=0.394), CVLT-II score (r=0.295, n=19, p=0.220/r=-0.048, n=19, p=0.845), DTI (r=0.247, n=19, p=0.308/r=0.040, n=19, p=0.870), or HSCT contrasts, Initiation>Repetition of 'REST' during Initiation (r=-0.038, n=19, p=0.877/r=-0.172, n=19, p=0.482) or Initiation>Cross-Fixation during Initiation (r=-0.187, n=19, p=0.443/r=0.099, n=19, p=0.686), nor; ii) successful FEP versus ARMS classifiers based either on CVLT-II score (r=0.052, n=15, p=0.853/r=-0.399, n=15, p=0.141), GM (r=0.208, n=15,

p=0.456/*r*=0.290, *n*=15, *p*=0.295) nor HSCT contrast Initiation>Cross-Fixation during Initiation (*r*=0.093, *n*=15, *p*=0.741/*r*=-0.118, *n*=15, *p*=0.676).

4.4 Discussion

Consistent with my first hypothesis based on the collective results of past univariate studies, FEP subjects were most readily discriminable from HCs with accurate classifiers generated by all modalities, with the exception of sMRI. In comparison, ARMS subjects were only distinguishable using sMRI and DTI data, but not fMRI, genetic, or, contrary to my hypothesis, neuropsychological data. With respect to my second hypothesis, using sMRI and neuropsychological data it was possible to directly discriminate between FEP and ARMS subjects at the individual level, though as hypothesised, not when using genetic data. Less consistent with my hypothesis, however, fMRI data was also able to successfully distinguish between the two groups at the single-subject level.

Taken together, there are number of potential inferences one may draw from the results in relation to FEP and the ARMS. Given that DTI data was able to successfully discriminate both FEP and ARMS subjects from HCs, for example, may be used to support the notion that patterns of WM alteration are associated with psychosis-risk; consistent with the results of previous univariate studies (Carletti, et al., 2012). Furthermore, the fact sMRI was able to discriminate ARMS from FEP and HC subjects may imply that patterns of GM alteration are specifically associated with an ARMS, but not, conversely, established psychosis. However, given previous reports of widespread significant effects (Shepherd et al. 2012), the absence of GM alteration in FEP subjects relative to HCs is surprising. One speculative explanation for this is that the FEP

subjects recruited here were less symptomatic than those of previous studies, potentially resulting in less severe alteration. This is supported by their relatively stable symptom profile (see Table 3.1 in section 3.2), which may possibly in turn, have been driven by downstream effects of exposure to anti-psychotic medication. This interpretation is made with caution however since the precise effects of such exposure remain unclear (Navari et al 2009), and since a successful classifier was not generated, this could also not be investigated quantitatively using a correlation analysis. In comparison, the finding that fMRI and neuropsychological data could differentiate FEP from ARMS and HCs, but not ARMS from HC subjects, may suggest that patterns of alteration in these two domains are specifically associated with conversion to psychosis. This inference is also true for individual genotype data, which was able to discriminate only between those with a FEP and HC subjects.

Methodologically, the results consolidate the notion that multivariate techniques such as SVM may be better suited to the development of a real-world clinical diagnostic tool than standard mass-univariate methods. Although no focal abnormalities survived univariate threshold either for sMRI nor fMRI for example (see section 3.3.3 and 3.3.5), overall patterns of alteration in data from these two modalities were still able to successfully discriminate between subjects. Furthermore, the ability to accurately distinguish FEP from HC subjects using genetic data supports the notion that individuals who suffer a FEP may be genetically predisposed to transitioning (Kéri, et al., 2009). As a non-invasive, easily obtained and relatively cheap data type, it could potentially serve as a good basis for future diagnostic tools in conjunction with clinical assessment. Similarly, the fact that CVLT-II score was able to distinguish FEP from both ARMS and HC subjects may represent another non-invasive and relatively inexpensive tool to help inform identification of individuals with a FEP. With specific

regard to ARMS subjects in comparison, the fact that only sMRI and DTI were able to distinguish them from HCs, might suggest that they are associated with patterns of neuroanatomical alteration that may occur in the absence of similar genetic, neurofunctional or cognitive patterns of alteration, though gross focal abnormalities may still be evident in these same modalities (see section 3.3.6)**.** This aspect of the results therefore provides tentative support to the use of structural MRI and DTI as a clinical aid in identifying those at increased risk of psychosis, but may be limited by the associated costs and technical expertise involved.

It should be acknowledged, however, that in comparison to the few previous studies to have applied SVM to psychosis (see section 4.1.1), the accuracies found here discriminating FEP, ARMS and HC subjects were relatively modest. As in many univariate studies, this inconsistency may have arisen from a number of possible methodological differences which include, but are not restricted to, the assessment tools used to identify subjects with respect to the ARMS; the strength of the scanner and the acquisition sequence used for the collection of neuroimaging data; the data-processing pipeline used to construct features for input into SVM; and the choice of SVM parameter settings (Orrù et al., 2012, Caruana & Niculescu-Mizil, 2006). Indeed, it is perhaps worth noting that as a relatively novel application to the field of psychiatry, efforts to identify the optimal criteria necessary for accurate discrimination using SVM are currently on-going, of which this study in addition to those outlined above, represent some examples.

In the context of developing real-world diagnostic tools therefore, two notes of caution must be considered. Firstly, the eventual use of genetic, neuropsychological and multimodal neuroimaging data in clinical practice would ultimately require substantially greater levels of diagnostic accuracy than those found here. One avenue to achieving

this may lie in the integration of different types of data within the same SVM allowing information from one modality to inform that of another, for example, as used recently by Yang et al. to discriminate ChSz patients from HCs (Yang et al., 2010). It remains however that any future translational implementation of SVM must account for the fact that the impact of misclassifying someone ill as healthy, may be worse than misclassifying someone healthy as ill. As such, a classifier able to detect patients with excellent sensitivity, but healthy individuals with only good specificity, may be preferred to a classifier with excellent specificity but only good sensitivity. Secondly, it should be noticed that the application of SVM could only reach the same level of diagnostic accuracy as traditional methods of clinical assessment since the development of the classifier is based on the distinction between groups in the training data, which ultimately relies on traditional diagnostic methods. Nevertheless, such technology may help in a clinical setting by discriminating between those most difficult to categorise using traditional methods of assessment alone. Furthermore, it could also potentially be used in a forensic setting as an objective means of reducing controversy in evaluations of mental illness and minimizing errors in detecting malingering (Sartori et al., 2011).

4.4.1 Limitations

The study's primary limitation was that at time of scanning, the majority of FEP patients were medicated, and correspondingly, symptomatically stable. Although there was no evidence for a significant impact of medication, it is possible that anti-psychotic exposure, or even other variables not considered here, may still have contributed to the classification in an as yet undetectable way, potentially confounding the inference one can draw from the successful discriminator. It should also be acknowledged furthermore that the medication measures used as variables to detect potential confounds (i.e. total

dose, average dose) may not have fully captured the historical and cumulative effects of exposure to anti-psychotics which may, in comparison, be more severe. Consequently, it cannot be ruled out for example that the successful FEP classifiers were simply distinguishing subjects who have, or have not, been exposed to anti-psychotics. However, it remains that since the exact nature and extent of the effects of anti-psychotic medication on brain structure are not yet known (Navari 2009), this is an issue not specific to the current investigation, but is instead one that applies to the majority of studies of psychiatric patients. A second limitation is that, in the absence of an optimally established method for correcting non-independent comparisons, two types of correction intended for multiple independent comparisons were used, which may have resulted in an increased risk of type II error. However, in acknowledgment of this, uncorrected results were also reported for completeness. A third limitation, applicable to any study with access restricted to their own sample, is that as a single centre, cross sectional, study it was not possible to draw inference regarding the generalizability across different research centres for any of the successful classifiers, nor at this stage, make any prediction of subsequent progression within the ARMS group. Finally, it was not possible to make any inference regarding specific features, i.e. neuroanatomical regions, CVLT-II task components, single SNPs, or risk alleles, given that in each case the entirety of the data entered into the SVM was used to generate the classifier, with the contribution made by each feature to the classifier only relevant in the context of all features (Schölkopf & Smola, 2002).

4.4.2 Conclusion

To conclude, the evidence presented here demonstrates that subjects who have had a FEP can be identified at the individual level using a range of biological and cognitive

measures including genetic, DTI, fMRI and neuropsychological data. In contrast sMRI and DTI were the only modalities that allowed identification of those at increased risk of psychosis with significant accuracy. The results also show for the first time, that FEP and ARMS subjects can be directly differentiated using neuropsychological, sMRI and fMRI data. From a clinical perspective, the results provide preliminary support to the translational development of SVM as a clinically useful diagnostic aid, highlighting patterns of genetic, cognitive, neuroanatomical and neurofunctional alteration that could, in future, be potentially used to inform identification of those with sub-clinical symptomatology and recent converters. Nevertheless, it must be stressed that the eventual use of this approach in everyday clinical practice would ultimately require considerably greater levels of diagnostic accuracy than found in the present investigation, with the integration of data, as explored in the next chapter, representing one possible solution.

Chapter 5

An Empirical Comparison of Different Approaches for Combining Multi-Modal Neuroimaging data with Support Vector Machine

5.1 Introduction

The multivariate analysis, support vector machine (SVM), has become an increasingly used tool in the study of neurological and psychiatric disease (Orrù, et al., 2012). Coinciding with a growing demand for clinically translatable research (Borgwardt & Fusar-Poli, 2012; Matthews, et al., 2006), such increase has been underlined by two properties in particular, i) the ability to classify a previously unseen individual into a predefined (diagnostic) category, at the single subject level, and ii) the ability to perform classification using a variety of measures, including neuroimaging data, which is by its nature, objective (Brammer, 2009). When considering the ultimate development of SVM as a real-world clinical aid for psychiatric and neurological disease, however, it is arguable that greater, more consistent, levels of discriminative accuracy are required than those currently reported (Orrù, et al., 2012), or, indeed, presented in the previous chapter. One method proposed to achieve such enhancement is the integration of data from different modalities, such that, complementary information from each modality can be used to inform that of the others (Kittler, et al., 1998). This is based on the premise that algorithms generated using different types of data will base

their classifications on distinct patterns of alteration and also make distinct pattern misclassifications. By combining different classifiers within a single SVM therefore, or, alternatively, by creating an ensemble of multiple single modality SVMs, enhanced levels of accuracy are intended through the derivation of a consensus decision, as opposed to a single modality, single decision, classifier (Kittler, et al., 1998).

5.1.1 Previous Studies of SVM Integration from Neurology and Psychiatry

In this context, existing applications of SVM integration involving Alzheimer's patients, have generally reported encouraging, albeit modest, increases in predictive averaging ranging between 3% and 7% relative to the best single modality classification accuracy (BSMCA) (Fan, et al., 2008; Hinrichs, et al., 2011; Zhang, et al., 2011) (see section 1.4 for further detail). With specific reference to psychosis comparatively, only one recent study investigating ChSz has been published, in which the authors report that using an integrative approach they were able to classify patients from HCs with 87.25% accuracy (Yang, et al., 2010) representing an increase of approximately 5% relative to the BSMCA. Despite these promising results, together, these four studies employed only two methods, or variations thereof, for integrating data within SVM, namely, majority voting and multi-kernel learning. Though alternative methods are available, to date, no investigation has yet been conducted examining the relative efficacies of a range of distinct integrative methods to combine multimodal data within the same clinical sample. It therefore remains unclear as to which approach may provide the greatest classification increase and in what context.

5.1.2 Aims and Hypotheses

The purpose of the investigation outlined in the current chapter therefore, was to review four different approaches that can be used to integrate data from multiple sources, namely, 1) an un-weighted 'simple' sum of kernels (SK), 2) multi-kernel learning (MKL), 3) prediction averaging (AV), and 4) majority voting (MV). The reason these particular methods were chosen was twofold: i) they represent those most frequently used in the very limited psychiatric and neurological literature (Fan, et al., 2008; Hinrichs, et al., 2011; Yang, et al., 2010; Zhang, et al., 2011) and/or, ii) they represent examples of the most readily implementable and technically user-friendly approaches currently available. In order to empirically examine their potential to enhance classification accuracy relative to the BSMCA, each approach was then applied to the same data set and the outcome recorded. In addition to this, for three of the four methods, integration was performed using combinations of both two, and three, data types in order to investigate the impact made by the number of data types being combined on levels of integrated accuracy (2-way data combination was not performed for MV for reasons outlined in section 5.2.2.4).

The specific data set to which each integrative method was applied comprised the structural, diffusion tensor and functional neuroimaging data detailed in Chapter 4, where I examined the ability of each modality separately, to successfully classify FEP and ARMS subjects from HCs, and each other. In brief, the results reported there showed that in conjunction with SVM, structural MRI (sMRI) data was able to discriminate ARMS from HCs and FEP subjects with significant (p<0.05) accuracies of 68.42% and 76.67% respectively; diffusion tensor imaging (DTI) data was able discriminate both ARMS and FEP subjects from HCs with a 65.79% accuracy; and, functional MRI (fMRI) data was able to discriminate FEP subjects from ARMS subjects

and HCs with up to 68.42% and 73.33% accuracy respectively. Using this data, I therefore examined the ability of the four integrative methods outlined, to enhance classification accuracy based on information combined from three distinct neuroimaging modalities; sMRI, DTI, and fMRI (comprising one of two functional contrasts), in order to discriminate ARMS from HC subjects, FEP from HC subjects, and/or ARMS from FEP subjects, relative to the BSMCA for each diagnostic comparison.

Since the number of studies to have applied integrative techniques to neuroimaging data is at present extremely limited (Hinrichs, et al., 2011; Yang, et al., 2010; Zhang, et al., 2011), for the purposes of the current study, my hypotheses regarding which method may work best were partly informed by the results of similar work in the field of proteomics. Specifically, in previous work conducted by Lewis and colleagues, the group applied SK, MKL, and single data type SVMs to the prediction of protein interactions and compared the results (Lewis, et al., 2006). With respect to their findings, the key observations applicable to the current study were threefold: i) an SK approach may outperform the relatively more computationally complex MKL approach when only few data types are being combined, ii) MKL may outperform an SK approach when more, rather than fewer, data types are being combined, and iii) for some data combinations, the best classification accuracy may be that achieved using just one of the constituent base (i.e. single modality) classifiers only. Given these observations therefore, in addition to those drawn from the few previous neurological and psychiatric studies outlined (Hinrichs, et al., 2011; Yang, et al., 2010; Zhang, et al., 2011), my hypotheses for the current investigation were as follows:

i) Classification accuracies would increase when integrating different types of data compared to when considering each type of data on its own

ii) For combinations of two data types, averaging and a sum of kernels approach would perform more successfully than a relatively more sophisticated multi kernel learning approach.

iii) For combinations of three data types, multi kernel learning would perform more successfully than either a sum of kernels, averaging or majority voting approach.

iv) The ability of each integration technique to enhance classification would vary with respect to the diagnostic comparison to which it was applied.

5.2 Materials and Methods

5.2.1 SVM

As outlined in Chapters 1, 2 and 4, originally developed in the early 1990s (Cortes & Vapnik, 1995), and stemming from statistical learning theory (Vapnik, 1999), SVM is a multivariate pattern recognition algorithm well suited to binary group classification. Specifically, the SVM aims to learn a decision function that correctly predicts the class label (conventionally denoted by $\boldsymbol{y}$ = +1 or -1) for each data point, based on a set of training examples, that can then be used to predict the labels for a set of unseen testing examples (Burges, 1998). Under the linear kernel formulation employed in this thesis, a dot product similarity measure is used to represent data in a symmetric, positive, kernel matrix. In this feature space SVM can be used to linearly separate groups (i.e. classes) of individuals (e.g. FEP and HC subjects) at the single subject level. The linear SVM decision function (Eq1) can be written as the dot product between each data vector ($\boldsymbol{x}$) and a vector of predictive weights ($\boldsymbol{w}$), from which the predicted class label can be derived by taking the sign (positive = class 1, negative = class 2) of the decision function. The weight vector represents an optimal separating hyperplane (OSH) in the input (i.e. voxel) space and can be represented in terms of the most difficult data points to classify – referred to as support vectors. The optimal weight vector is determined by maximising the margin between groups thus aiming to ensure good generalisation to new data, an approximately unbiased estimate of which can be obtained using cross-validation (Hastie, 2001; Lemm, et al., 2011). The SVM objective function is provided in equations 2 and 3, reflecting the primal, and dual, space representations respectively. Here, $\boldsymbol{x_i}^T\boldsymbol{x_j}$ denotes the inner product between data samples (i.e. subjects), $\boldsymbol{b}$ denotes

offset, $\boldsymbol{\xi_i}$ denote slack variables which permit data to be misclassified in the training set, $\boldsymbol{\alpha_i}$ denote Lagrange multipliers and $\boldsymbol{C}$ is a parameter regulating the balance between maximising the margin between data points and allowing misclassification. For a more detailed description of the general framework of SVM please refer to works either by Burges (Burges, 1998) or Schölkopf and Smola (Schölkopf & Smola, 2002), or for an overview of SVM in the context of neuroimaging, Pereira et al. and/or Lemm et al. (Lemm, et al., 2011; Pereira et al., 2009).

$$\boldsymbol{f(x, w) = (w^T x + b)} \qquad \textbf{(1)}$$

$$\textbf{minimize}_{w,\xi} \quad \boldsymbol{\frac{1}{2}\|w\|^2 + C\sum_i \xi_i} \qquad \textbf{(2)}$$

$$\textbf{subject to (s.t .)} \quad \boldsymbol{y_i\left(w^T x_i + b\right) \geq 1 - \xi_i}$$

$$\boldsymbol{\xi_i \geq 0 \quad \forall i}$$

$$\textbf{maximize}_{\alpha} \quad \boldsymbol{\sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j {x_i}^T x_j} \qquad \textbf{(3)}$$

$$\textbf{s.t.} \quad \boldsymbol{\sum_i \alpha_i y_i = 0}$$

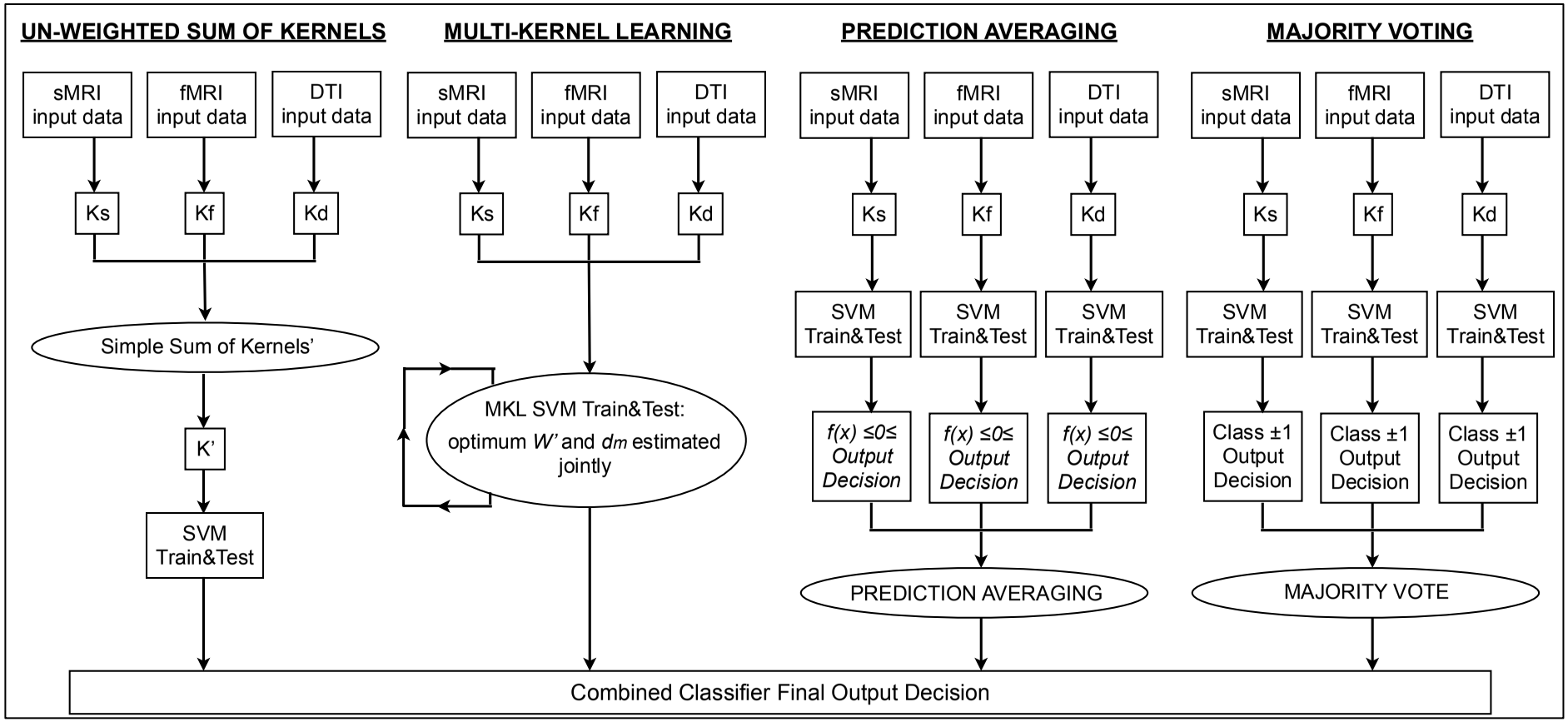
$$\boldsymbol{0 \leq \alpha_i \leq C \qquad \forall i}$$

Equations 2 and 3 represent convex optimisation problems and can be efficiently optimised with conventional quadratic solvers, providing the optimum solution to equation 1. In the present work, the libSVM implementation was employed (Chang & Lin, 2011), as implemented in the PROBID software toolbox (www.brainmap.co.uk/probid.htm), with the value of the SVM regularisation parameter $\boldsymbol{C}$ fixed at one (default value).

5.2.2 Combining classifiers

As briefly outlined in Chapters 1 and 2, in order to generate a single output classification decision based on data integrated from multiple sources, two potential options are, 1) find a linear combination of the kernel matrices representing each data modality, in order to train and test a single SVM based on a combination of all input data; in which case, the weight vectors from each data type are estimated jointly, or 2) train multiple single modality classifiers and subsequently combine the output decisions to generate a single ensemble decision function; in which case, the weight vectors for each data type are estimated independently. Of these two options, an un-weighted sum of kernels and a multi-kernel learning approach are variations of option 1, whilst prediction averaging and majority voting are variations of option 2. In Fig. 5.1 a representative pipeline is depicted showing the steps used for each integrative approach. As detailed in Chapter 4, on the proviso that equal subject numbers are used for each diagnostic group being compared, for each (integrated) classifier a balanced accuracy can be calculated from the mean of the (integrated) specificity (i.e. number of class 1 subjects, e.g. patients, correctly classified as class 1) and (integrated) sensitivity (i.e. the number of class 2 subjects, e.g. HCs, correctly classified as class 2) (see Box 1 in section 4.2.3.6.3).

Figure 5.1. Flowchart depicting the processing pipeline for each type of integrative approach



MRI: magnetic resonance imaging, sMRI: structural MRI, DTI: diffusion tensor imaging, fMRI: functional MRI, Ks/f/d: kernel matrix for sMRI/fMRI/DTI data, SVM: support vector machine, K': integrated kernel matrix, f(x): SVM decision function, *W'*: optimum vector of predictive weights obtained using K', d_m: optimum weight coefficient assigned to each base kernel.

5.2.2.1 Un-weighted Simple Sum of Kernels

A well-known property of kernels is that they can be combined via linear operations (e.g. addition and multiplication) to yield a valid kernel. As described, a linear kernel matrix was used in the present work to represent the similarity between data points (i.e. samples) within each data modality (Eq4). Thus, a simple way to combine data modalities is simply to add the kernel matrices. Importantly, different modalities may have different numbers of features and may also be scaled differently. To account for this, it was necessary for each kernel to be first normalised, before being summed together to create a new kernel matrix representing data from all modalities (Eq5: example shown represents combining two data types only, e.g. kernel matrix for data type 1: $\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j})$, and for data type 2: $\boldsymbol{K}(\boldsymbol{x_i}', \boldsymbol{x_j}')$). This equation is equivalent to first dividing each data vector by its norm (i.e. its length), then concatenating the feature vectors for all modalities. Under this framework, the data from each source is assigned an equal weighting in terms of its contribution in defining the OSH. A SVM is then trained and tested (Eq2 and 3) using this new integrated kernel (***K'***), such that any classification decision is based on a combination of data from each integrated source.

$$\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \langle \boldsymbol{x_i} . \boldsymbol{x_j} \rangle \tag{4}$$

$$\boldsymbol{K'} = \frac{\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j})}{\sqrt{\left(\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_i})\boldsymbol{K}(\boldsymbol{x_j}, \boldsymbol{x_j})\right)}} + \frac{\boldsymbol{K}(\boldsymbol{x_i}', \boldsymbol{x_j}')}{\sqrt{\left(\boldsymbol{K}(\boldsymbol{x_i}', \boldsymbol{x_i}')\boldsymbol{K}(\boldsymbol{x_j}', \boldsymbol{x_j}')\right)}} \tag{5}$$

5.2.2.2 Multi-Kernel Learning

Consistent with an SK approach, the integration step of MKL involves direct manipulation of the kernel matrices representing each set of input data, such that a single kernel simultaneously representing all data being combined is used to train and test a single integrated SVM. In comparison to SK however, MKL employs an additional phase of 'learning', which involves iteratively weighting each single modality kernel based on its relative contribution to the classification. This is intended to alter the dimension of each kernel, reducing the influence of 'noisy' data types whilst still utilising any complementary information it has to offer. For example, in a feature space based on an integration comprising two kernels ($\boldsymbol{K_1}$ and $\boldsymbol{K_2}$), in which the dimension of the first kernel has been scaled by a factor of 3, the integrated kernel, ($\boldsymbol{K'}$) would be represented thus:

$$\boldsymbol{K' = 3K_1 + K_2} \qquad \textbf{(6)}$$

Currently, a number of optimization approaches exist which allow the optimal individual kernel weights and SVM parameters to be computed simultaneously, when the kernel matrices used are semi-definite positive (Lanckriet et al., 2004; Sonnenburg et al., 2006). Of those, the particular technique employed here is called SimpleMKL (Rakotomamonjy et al., 2008). Notably, this approach combines the iterative weighting of base kernels and SVM training and testing using a single objective function, such that the optimum kernel weights ($\boldsymbol{d_m}$) and vector of predictive feature weights ($\boldsymbol{w}$) (based on all data combined) are estimated jointly (see Fig. 5.1). This means the kernel upon which the SVM is trained is a linear combination of optimally weighted base kernels (Eq6), more formally expressed in equation 7. Here, $\boldsymbol{M}$ denotes the total number

of kernels ($\boldsymbol{m} = 1, \ldots, \boldsymbol{M}$), $\boldsymbol{K_m(x_i, x_j)}$ denotes the inner product between two data points for each base kernel, and $\boldsymbol{d_m}$ denotes the weight assigned to each base kernel.

$$\boldsymbol{K'(x_i, x_j) = \sum_{m=1}^{M} d_m K_m(x_i, x_j)} \qquad \textbf{(7)}$$

$$\textbf{s.t.} \qquad \boldsymbol{d_m \geq 0}, \quad \textbf{and} \quad \boldsymbol{\sum_{m=1}^{M} d_m = 1}$$

Unlike a single data, or SK, SVM however, in order to solve the problem in dual space (cf. Eq3), the Simple MKL SVM objective function in the primal space must be first transformed into a second constrained optimization problem, from which a more easily soluble dual form can be obtained. This is provided in equations 8 and 9 reflecting the primal, and dual, space representations respectively. Here, $\boldsymbol{f_m(x_i)}$ denotes the decision function for the $\boldsymbol{i_{\text{th}}}$ sample based on the $\boldsymbol{m_{th}}$ kernel, ***J(d)*** denotes the optimal SVM objective value, $\boldsymbol{\|f_m\|^2_{H_m}}$ (equivalent to $\boldsymbol{\|w\|^2}$ in Eq2) denotes the regularization term with respect to each feature space ($\boldsymbol{H_m}$), and the remaining terms are as above.

$$\textbf{minimize}_{\boldsymbol{d}} \qquad \boldsymbol{J(d)} \quad \textbf{such that} \quad \boldsymbol{\sum_{m=1}^{M} d_m = 1,\ d_m \geq 0} \qquad \textbf{(8)}$$

$$\textbf{s.t.} \qquad \boldsymbol{J(d)} = \begin{cases} \boldsymbol{\min_{\{f\},b,\xi} \frac{1}{2} \sum_m \frac{1}{d_m} \|f_m\|^2_{H_m} + C \sum_i \xi_i \quad \forall i} \\ \boldsymbol{s.t. \quad y_i \sum_m^m f_m(x_i) + y_i b \geq 1 - \xi_i} \\ \boldsymbol{\xi_i \geq 0 \quad \forall i} \end{cases}$$

$$\textbf{maximize}_{\boldsymbol{\alpha}} \qquad \boldsymbol{\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_m d_m K_m(x_i, x_j)} \qquad \textbf{(9)}$$

$$\textbf{s.t.} \qquad \boldsymbol{\sum_i \alpha_i y_i = 0}$$

$$\boldsymbol{0 \leq \alpha_i \leq C \qquad \forall i}$$

As convex optimization problems, equations 8 and 9 can be efficiently optimised with conventional quadratic solvers. This process is then iteratively repeated each time leading to a reduction in the objective function (Eq8), until the duality gap (i.e. the difference between primal and dual objective values) approximates to zero, at which point the optimal conditions for the kernel weight coefficients and SVM parameters have been met. For a more detailed description of the framework used for SimpleMKL please refer to Rakotomamonjy et al. (2008).

5.2.2.3 Prediction Averaging

In contrast to SK, or MKL, AV integrates modalities at the level of predictions (i.e. forming an ensemble decision after each base classifier has been trained and tested on a single modality; see Fig. 5.1). Rather than using the binary class labels (i.e. ±1) however, integration is achieved by taking the mean of the predictive function values over all modalities and computing its sign to derive an average class prediction (Eq10). Hence for a sample $\boldsymbol{x}$, classified by $\boldsymbol{M}$ ($\boldsymbol{m}$ = 1, … , $\boldsymbol{M}$) individual classifiers, each represented by a decision function $\boldsymbol{f_m(x)}$ (Eq3), the final class based on integrated data using AV can be predicted by:

$$\boldsymbol{x_{class} = sgn\left(\frac{1}{M}\sum_{m=1}^{M} f_m(x)\right)} \qquad \textbf{(10)}$$

5.2.2.4 Majority Voting

Consistent with AV, MV also performs integration at the level of the predictions. However, for MV only the sign (i.e. binary outcome) of the decision function is considered, rather than its absolute value as in AV. Under the MV approach, the final class label is therefore determined by assigning the sample to the class obtaining the largest number of predictions amongst the base classifiers.

Since MV only relies on the binary outcome, in cases where an even number of data types are combined, it is possible that tied decisions may occur, in which case, they must be broken using a heuristic chosen *a priori*. In the current investigation MV was therefore not performed for data combined from two modalities, given the likelihood that the integrated output classification would be overly influenced by the chosen heuristic, used when the two individual modality classifiers made opposing classifications.

5.2.3 Data used for SVM Integration

In order to be able to integrate different modalities, the very same pairs of subjects must be used across different modalities; however while the number of pairs used for sMRI, DTI and fMRI data were exactly the same, a smaller number of pairs were used for the independent investigation of genetic and cognitive modalities as reported in Chapter 4. In order to use the greatest possible number of pairs, and therefore maximize the reliability of the inferences, I therefore focused on sMRI, DTI and fMRI data for the purpose of comparing the different integration approaches. Of the six fMRI contrasts tested in the previous chapter furthermore, the two selected for inclusion here were chosen on the basis that the conditions being contrasted represent the most cognitively

divergent of the six available, and were therefore most likely to result in the greatest activation differences, offering the maximum chance for successful integrative classification. These contrasts were, i) generation of an overt verbal initiation response > visual cross-fixation during the initiation condition (In>CFI), and ii) generation of an overt verbal suppression response > visual cross-fixation during the suppression condition (Su>CFS).

5.2.4 SVM Integration: an empirical comparison

In order to obtain a general measure representing the relative ability of each technique to enhance classification accuracy based on the integration of data from different modalities overall, I performed a non-parametric Wilcoxon test comparing the integrated accuracies achieved by each method against every other method, collapsed across binary diagnostic comparisons, both for two and three data type combinations. The results of this test are presented in figures 5.2 and 5.3, alongside graphic visualisations showing the relative difference between the classification accuracy achieved by each integrative method and the BSMCA, for each diagnostic contrast, for combinations of both two (Fig. 5.2) and three (Fig. 5.3) data types.

5.3. Results

5.3.1 Un-weighted Sum of Kernels

5.3.1.1 Data Integrated from 2 modalities

Using SK, the ability of sMRI and DTI data combined to differentiate FEP from HC subjects was increased to 71.05% representing an approximate increase of 6% relative to the BSMCA. Furthermore, combining DTI and fMRI data using an SK approach, it was possible to discriminate FEP from ARMS subjects with 83.33% accuracy representing an approximate increase of 10% relative to the BSMCA (see Fig. 5.2 and Table 5.1).

5.3.1.2 Data Integrated from 3 modalities

By combining three different data types using SK it was not possible to enhance classification accuracy relative to the BSCMA for any of the diagnostic comparisons.

Table 5.1. Classification accuracies combining sMRI, DTI and fMRI data in two-, and three-, way combinations using an un-weighted sum of kernels to discriminate ARMS from HCs, FEP from HCs and FEP from ARMS subjects.

Data Combination	**Un-weighted Sum of Kernels**		
	ARMS x HC	**FEP x HC**	**FEP x ARMS**
GM + FAS (%)	63.16 (68.42 + 65.79)	71.05 * (63.16+65.79)	76.67 (76.67+56.67)
GM + In>CFI (%)	47.37 (68.42 + 36.84)	65.79 (63.16+68.42)	63.33 (76.67+73.33)
GM + Su>CFS (%)	44.74 (68.42 + 60.53)	60.53 (63.16+47.37)	53.33 (76.67+53.33)
FAS + In>CFI (%)	52.63 (65.79+36.84)	63.16 (65.79+68.42)	83.33 ** (56.67+73.33)
FAS + Su>CFS (%)	63.16 (65.79+60.53)	55.26 (65.79+47.37)	56.67 (56.67+53.33)
GM + FAS + In>CFI (%)	50.00 (68.42+65.79+36.84)	63.16 (63.16+65.79+68.42)	76.67 (76.67+56.67+73.33)
GM + FAS + Su>CFS (%)	60.53 (68.42+65.79+60.53)	55.26 (63.16+65.79+47.37)	63.33 (76.67+56.67+53.33)

sMRI: structural MRI, DTI: diffusion tensor imaging, fMRI: functional MRI, GM: grey matter, FAS: fractional anisotropy skeleton, In: Initiation, CFI: Cross-fixation during Initiation, Su: Suppression, CFS: Cross-fixation during Suppression, ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Subjects. Starred figures represent accuracy increases of 0<10% (*), or ≥10% (**), relative to the single best accuracy of the modalities used in the integration. Figures in brackets are the accuracies for the single modalities ordered as per the data combination column.

5.3.2 Multi Kernel Learning

5.3.2.1 Data Integrated from 2 modalities

Based on the integration of two data types, MKL was unable to enhance classification accuracy for any of the diagnostic comparisons relative to the BSMCA (see Fig. 5.2 and Table 5.2).

5.3.2.2 Data Integrated from 3 modalities

Combining sMRI, DTI and fMRI (contrast: Suppression>Cross-Fixation during Suppression) data using MKL, the ability to distinguish FEP from HC subjects was increased to 71.05% representing an approximate increase of 6% relative to the BSMCA (see Fig. 5.3 and Table 5.2).

Table 5.2. Classification accuracies combining sMRI, DTI and fMRI data in two-, and three-, way combinations using multi kernel learning to discriminate ARMS from HCs, FEP from HCs and FEP from ARMS subjects.

Data Combination	**Multi Kernel Learning**		
	ARMS x HC	**FEP x HC**	**FEP x ARMS**
GM + FAS (%)	60.53 (68.42 + 65.79)	55.26 (63.16+65.79)	73.34 (76.67+56.67)
GM + In>CFI (%)	50.00 (68.42 + 36.84)	52.63 (63.16+68.42)	50.00 (76.67+73.33)
GM + Su>CFS (%)	52.63 (68.42 + 60.53)	52.63 (63.16+47.37)	50.00 (76.67+53.33)
FAS + In>CFI (%)	53.95 (65.79+36.84)	50.00 (65.79+68.42)	50.00 (56.67+73.33)
FAS + Su>CFS (%)	50.00 (65.79+60.53)	50.00 (65.79+47.37)	50.00 (56.67+53.33)
GM + FAS + In>CFI (%)	44.74 (68.42+65.79+36.84)	63.16 (63.16+65.79+68.42)	70.00 (76.67+56.67+73.33)
GM + FAS + Su>CFS (%)	44.74 (68.42+65.79+60.53)	71.05 * (63.16+65.79+47.37)	63.33 (76.67+56.67+53.33)

sMRI: structural MRI, DTI: diffusion tensor imaging, fMRI: functional MRI, GM: grey matter, FAS: fractional anisotropy skeleton, In: Initiation, CFI: Cross-fixation during Initiation, Su: Suppression, CFS: Cross-fixation during Suppression, ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Subjects. Starred figures represent accuracy increases of 0<10% (*), or ≥10% (**), relative to the single best accuracy of the modalities used in the integration. Figures in brackets are the accuracies for the single modalities ordered as per the data combination column.

5.3.3 Prediction Averaging

5.3.3.1 Data Integrated from 2 modalities

Using AV, the ability of sMRI and DTI data combined to differentiate ARMS from HC subjects was increased to 71.05% representing an approximate increase of 3% relative to the BSMCA. Similarly, combining sMRI and fMRI (contrast: Initiation>Cross-Fixation during Initiation) data using AV enhanced classification of FEP from HC subjects to 71.05%, representing an approximate increase of 3% relative to the BSMCA. When applied to the integration of DTI with fMRI data in order to discriminate FEP from ARMS subjects, AV was able to increase classification accuracy to 86.67% and 66.67%, using the Initiation>Cross-Fixation during Initiation and Suppression>Cross-Fixation during Suppression contrasts respectively, representing approximate increases of 13% and 10% relative to the BSMCA in each case (see Fig. 5.2, and Table 5.3).

5.3.3.2 Data Integrated from 3 modalities

Combining sMRI, DTI and fMRI (contrast: Initiation>Cross-Fixation during Initiation) data using AV, the ability to distinguish FEP from HC subjects, and FEP from ARMS subjects, was increased to 71.05% and 83.33% respectively. In each case this represented an approximate increase of 3% and 7% relative to the BSMCA (see Fig. 5.3 and Table 5.3).

Table 5.3. SVM classification accuracies combining sMRI, DTI and fMRI data in two-, and three-, way combinations using prediction averaging to discriminate ARMS from HCs, FEP from HCs and FEP from ARMS subjects.

Data Combination	**Prediction Averaging**		
	ARMS x HC	**FEP x HC**	**FEP x ARMS**
GM + FAS (%)	71.05 * (68.42 + 65.79)	60.53 (63.16+65.79)	70.00 (76.67+56.67)
GM + In>CFI (%)	47.37 (68.42 + 36.84)	71.05 * (63.16+68.42)	76.67 (76.67+73.33)
GM + Su>CFS (%)	63.16 (68.42 + 60.53)	57.89 (63.16+47.37)	66.67 (76.67+53.33)
FAS + In>CFI (%)	52.63 (65.79+36.84)	65.79 (65.79+68.42)	86.67 ** (56.67+73.33)
FAS + Su>CFS (%)	65.79 (65.79+60.53)	50.00 (65.79+47.37)	66.67 ** (56.67+53.33)
GM + FAS + In>CFI (%)	57.89 (68.42+65.79+36.84)	71.05 * (63.16+65.79+68.42)	83.33 * (76.67+56.67+73.33)
GM + FAS + Su>CFS (%)	65.79 (68.42+65.79+60.53)	57.89 (63.16+65.79+47.37)	66.67 (76.67+56.67+53.33)

sMRI: structural MRI, DTI: diffusion tensor imaging, fMRI: functional MRI, GM: grey matter, FAS: fractional anisotropy skeleton, In: Initiation, CFI: Cross-fixation during Initiation, Su: Suppression, CFS: Cross-fixation during Suppression, ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Subjects. Starred figures represent accuracy increases of 0<10% (*), or ≥10% (**), relative to the single best accuracy of the modalities used in the integration. Figures in brackets are the accuracies for the single modalities ordered as per the data combination column.

5.3.4 Majority Voting

5.3.4.1 Data Integrated from 3 modalities

Using MV, it was possible to discriminate FEP from HC subjects with 71.05% accuracy based on the three-way combination of sMRI, DTI and fMRI (contrast: Initiation>Cross-Fixation during Initiation) data. This represented an approximate increase of 3% relative to the BSMCA (see Fig. 5.3 and Table 5.3).

Table 5.4. SVM classification accuracies combining sMRI, DTI and fMRI data in a three way combination using majority voting to discriminate ARMS from HCs, FEP from HCs and FEP from ARMS subjects.

Data Combination	**Majority Voting**		
	ARMS x HC	**FEP x HC**	**FEP x ARMS**
GM + FAS + In>CFI (%)	57.89 (68.42+65.79+36.84)	71.05 * (63.16+65.79+68.42)	73.33 (76.67+56.67+73.33)
GM + FAS + Su>CFS (%)	65.79 (68.42+65.79+60.53)	63.16 (63.16+65.79+47.37)	63.33 (76.67+56.67+53.33)

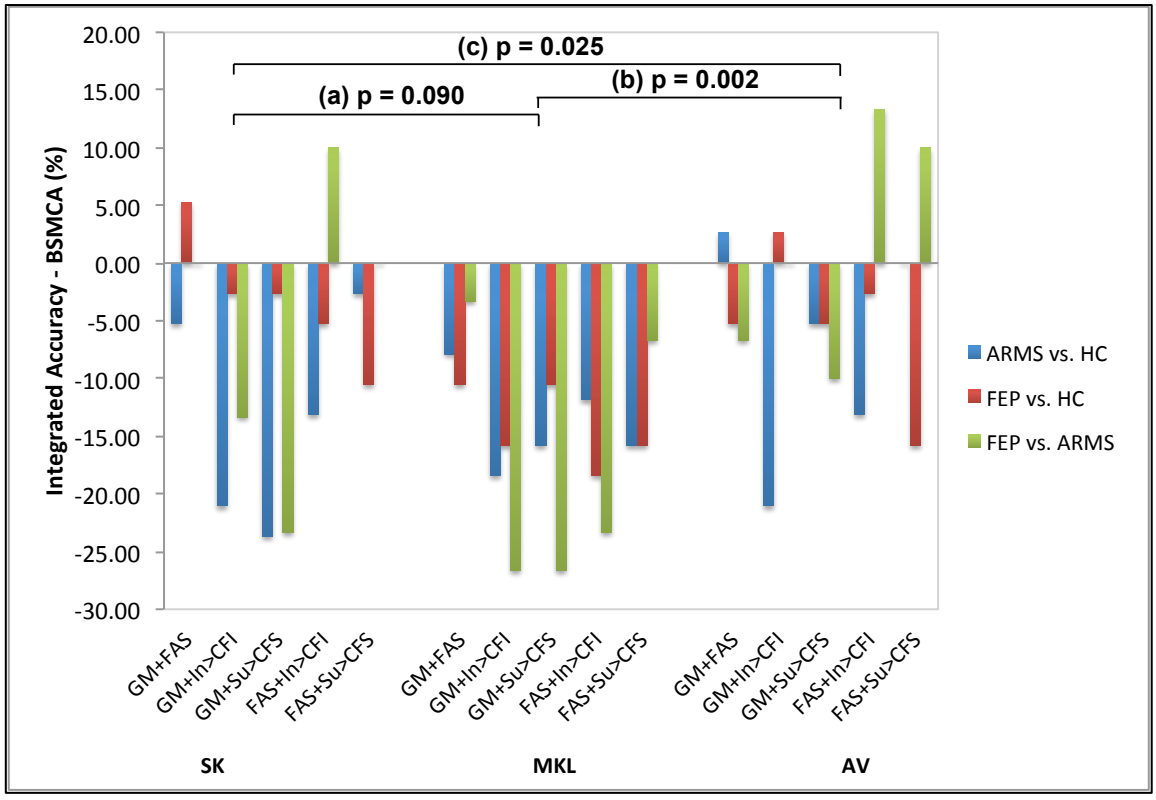
sMRI: structural MRI, DTI: diffusion tensor imaging, fMRI: functional MRI, GM: grey matter, FAS: fractional anisotropy skeleton, In: Initiation, CFI: Cross-fixation during Initiation, Su: Suppression, CFS: Cross-fixation during Suppression, ARMS: At-Risk Mental State, FEP: First Episode Psychosis, HC: Healthy Subjects. Starred figures represent accuracy increases of 0<10% (*), or ≥10% (**), relative to the single best accuracy of the modalities used in the integration. Figures in brackets are the accuracies for the single modalities ordered as per the data combination column.

5.3.5 An empirical comparison of methods

5.3.5.1 Data combined from two modalities

With respect to a two modality combination, when collapsed across all SVM diagnostic comparisons the results of the Wilcoxon test comparing the integrated accuracy achieved by each method, for each two way combination, gave the following best-to-worst ranking of methods based on their respective p-values: prediction averaging, simple sum of kernels, multi-kernel learning (see Fig. 5.2) (MV was not performed for data combined from two modalities due to the potential over influence of the heuristic used where ties occur. See section 5.2.2.4 for detail). These results are consistent with the best integrated accuracies achieved by each method with respect to each individual diagnostic comparison, with MKL broadly outperformed by both AV and SK (see Fig. 5.2). Across diagnostic comparisons furthermore, greater integrated accuracies were achieved for the FEP versus ARMS comparison by each of the three integrative approaches, relative to the other two diagnostic comparisons.

Figure 5.2. Difference between the integrated accuracy achieved using SK, MKL, or AV, and the BSMCA, discriminating ARMS and FEP subjects from HCs, and each other, using two-way combinations of sMRI, DTI and fMRI data.

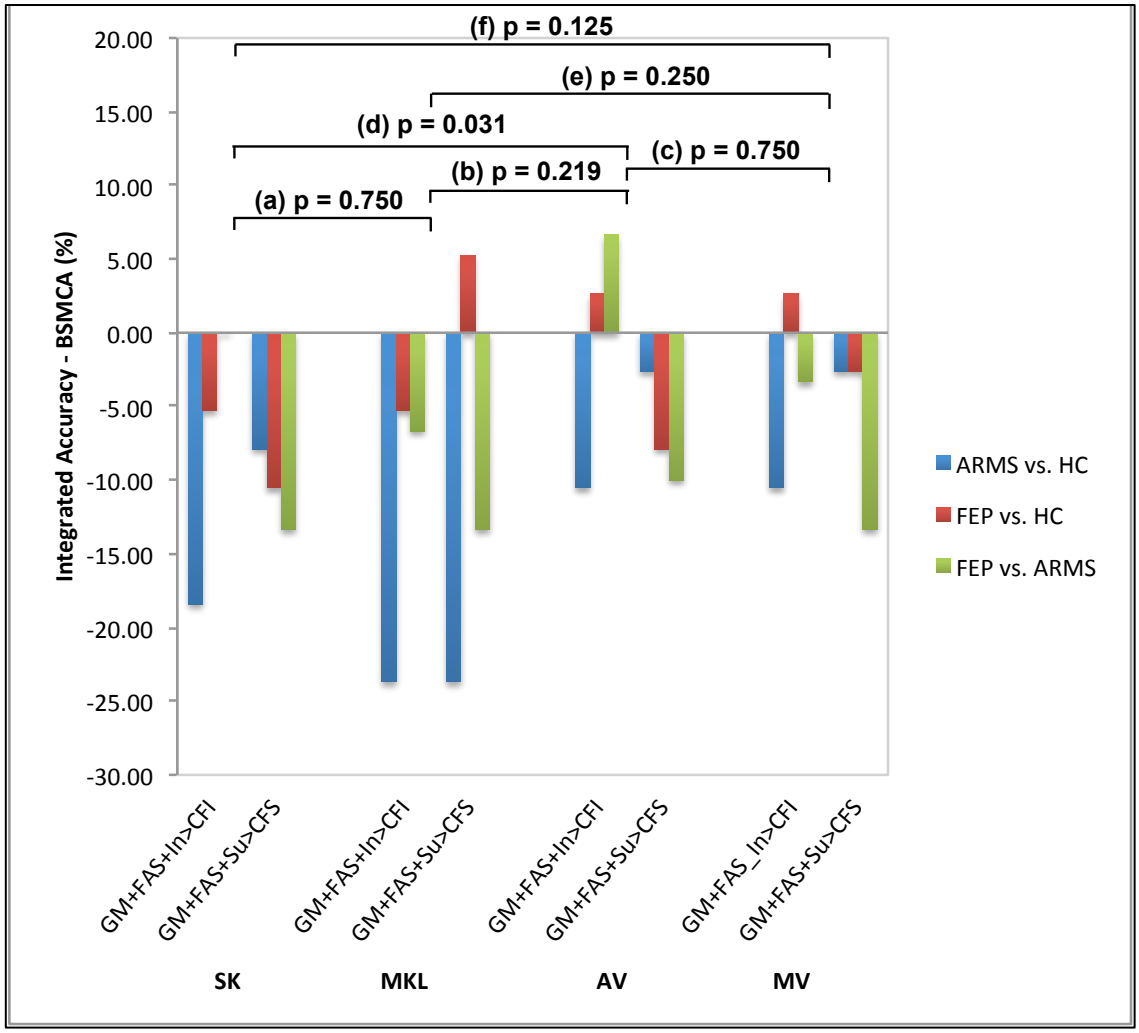


SK: Un-weighted simple sum of kernels, MKL; Multi-Kernel Learning, AV: Prediction Averaging. GM: grey matter, FAS: fractional anisotropy skeleton, In>CFI: generation of an overt verbal initiation response > visual cross-fixation during the initiation condition, Su>CFS: generation of an overt verbal suppression response > visual cross-fixation during the suppression condition, BSMCA: best single modality classification accuracy. P-values refer to the results of a Wilcoxon test comparing the integrated classification accuracies, collapsed across diagnostic comparisons, of; (a) SK and MKL, (b) AV and MKL, and (c) AV and SK.

5.3.5.2 Data combined from three modalities

With respect to a three modality combination, when collapsed across all SVM diagnostic comparisons the results of the Wilcoxon test comparing the integrated accuracy achieved by each method, for each three way combination, though not all significant, gave a similar best-to-worst ranking of methods based on their respective p-values: prediction averaging, majority voting, simple sum of kernels and multi-kernel learning. As shown in Fig. 5.3, these results are broadly consistent with the best integrated accuracy achieved by each method with respect to each individual diagnostic comparison, with AV performing better than MV, which in turn performed better than SK and MKL, which were relatively equal. Across diagnostic comparisons furthermore, consistent with the two-way combinations, the best integrated accuracies were generally achieved for the FEP versus ARMS comparison, followed by the FEP versus HC, and then ARMS versus HC comparisons.

Figure 5.3. Difference between the integrated accuracy achieved using SK, MKL, AV, or MV, and the BSMCA, discriminating ARMS and FEP subjects from HCs, and each other, using three-way combinations of sMRI, DTI and fMRI data



SK: Un-weighted simple sum of kernels, MK; Multi-Kernel Learning, AV: Prediction Averaging, MV: Majority Voting. GM: grey matter, FAS: fractional anisotropy skeleton, In>CFI: generation of an overt verbal initiation response > visual cross-fixation during the initiation condition, Su>CFS: generation of an overt verbal suppression response > visual cross-fixation during the suppression condition, BSMCA: best single modality classification accuracy. P-values refer to the results of a Wilcoxon test comparing the integrated classification accuracies, collapsed across diagnostic comparisons, of; (a) SK and MKL, (b) AV and MKL, (c) AV and MV, (d) AV and SK, (e) MV and MKL, and (f) MV and SK

5.4. Discussion

In the current study I employed a comparative approach to investigate the relative abilities of four distinct integrative methods, to enhance the classification accuracy of a SVM analysis by combining multimodal data. Specifically, each method was applied to three separate diagnostic comparisons comprising FEP, ARMS and HC subjects, utilising combinations of data from three distinct neuroimaging modalities sMRI, fMRI and/or DTI. With respect to the first of my hypotheses, the results provided a mixed picture. Consistent with the hypothesis, I found classification improvements of up to 13% relative to the BSMCA were made possible using an integrative approach, which, even in relation to previous studies (Hinrichs, et al., 2011; Yang, et al., 2010; Zhang, et al., 2011), represents a substantial improvement. Contrary to my hypothesis however, such improvements were made in the context of considerable variation in the levels of integrated accuracy achieved overall, with the optimum classification frequently represented by the BSMCA (see figures 5.2 and 5.3). With respect to my second hypothesis, the results showed that in the context of data integration from two modalities alone, an un-weighted simple sum of kernels did perform better than a relatively more sophisticated, weighted, multi kernel learning approach. However, the results also showed AV to be a more successful integrative technique than either SK or MKL. With respect to my third hypothesis, the results showed that for data combined from three modalities overall, MKL did outperform SK in terms of the number and magnitude of classification increases, but was in turn outperformed by AV based on the same criteria. In addition, a Wilcoxon test showed that when collapsed across diagnostic comparisons, MV performed equally well as SK and MKL, with only the AV results statistically greater than those of SK (see Fig. 5.3). In agreement with my fourth hypothesis, the data suggests that the ability of each integrative approach to increase

classification accuracy is largely dependent on the diagnostic comparison to which it is applied. This is aptly demonstrated in figures 5.2 and 5.3, which show that in absolute terms the most classification improvements were made for the FEP versus ARMS comparison, followed by the FEP versus HC comparison, and finally the ARMS versus HC comparison.

Taken together, the results suggest that whilst the integration of different data types can potentially yield relatively substantial increases in classification accuracy, the frequency of such instances may be limited. Furthermore, the findings highlight that the potential of each integrative method to enhance classification accuracy appears to be: i) differentially suited to different diagnostic comparisons, ii) influenced by the number of different data types being integrated, and iii) influenced by the specific type of data being integrated.

With respect to the influence of diagnostic comparison, as shown by figures 5.2 and 5.3 there is a clear distinction between each of the four integrative methods to enhance classification accuracy dependent on the diagnostic comparison to which it is applied. One possible explanation for this is that the different data types are indexing different manifestations of common underlying factors (such as genetic or cellular aetiology) to a greater, or lesser degree, in some diagnostic comparisons relative to others. In cases where this is to a greater degree, the process of integration would be unlikely to provide enhanced classification since the different single modality classifiers would ultimately be driven by similar patterns of underlying alteration, with little complementary information therefore available to be shared between them (i.e. two different classifiers which make the same predictions for the same subjects have little complementary information to add to one another).

With respect to the second factor, the results suggest that the assumption that greater numbers will necessarily result in greater accuracy may not be true. For example, the integrations with greatest increase were based on data combined from two, rather than three, modalities (see figures 5.2 and 5.3). The results also support the notion that when only few data types of approximately equal weight are being integrated, less computationally complex techniques such as prediction averaging and a simple summing of kernels may provide comparable, if not greater, levels of integrated accuracy in comparison to more computationally complex approaches such as MKL. This is consistent with Lewis et al. who reported similar findings based on their application of integrated SVM techniques to protein interaction prediction (Lewis, et al., 2006). This concept is further supported by the results of other recent studies, which report that MKL may instead be the preferred option when data from tens, or hundreds, of different sources are being combined relative to just two or three. For example, a study by Damoulas et al., also investigating protein interactions, demonstrated that a single integrated MKL classifier was able to improve classification accuracy by approximately 8%, relative to the BSMCA (Damoulas & Girolami, 2008), when the data from which it was derived, and the classes being classified, would have represented over 2000 single binary SVM classifiers. One proposed benefit of MKL still partially emergent in the data here however, is the explicit ability to down-regulate the weight of a 'noisy' data set, whilst still utilising its complementary information to achieve enhanced classification. For example, MKL was able to combine the fMRI contrast, Su>CFS, which by itself had only been able to classify FEP from HC subjects with an insignificant 47.37% accuracy, with sMRI and DTI data, and enhance overall classification accuracy by 6% relative to the BSMCA (see table 5.2 and figure 5.3). In comparison, using the *same* three-way data combination, the remaining integrative

techniques were unable to enhance classification accuracy, possibly due to the un-weighted contribution of the fMRI contrast acting as 'noisy' interference.

In addition to the number of modalities, it also seems evident that the specific types of modality being integrated is a third important factor. For example, whilst the combination of DTI and fMRI data using SK and AV provided an approximate increase 10% and 13% respectively with regard to the differentiation of FEP and ARMS subjects, combining sMRI with DTI data, or the same fMRI contrast, using the same integrative methods, did not result in a similarly increased classification accuracy (see Tables 5.1 and 5.2). As above, this too may be driven by the possibility that different data types index manifestations of illness which are more, or less, closely linked to common underlying aetiology, therefore varying the amount of complementary information one data type can offer another. As such, when considering SVM as a research, and potential real-world clinical, tool, it should be emphasised that the ability to classify different clinical groups with the best accuracy will be associated with specific data types which should be clarified.

In summary, the results here support the notion that the integration of different data types has, in principle, the potential to enhance the differentiation of individuals from distinct clinical groups both from HCs and each other. However, in considering the clinical development of SVM underlined by an integrative approach, two notes of caution must be observed. Firstly, as tables 5.1-5.4 show, though integration of data has the potential to provide a substantial increase in classification accuracy of up to 13%, there are many instances in which the optimal classification accuracy differentiating two groups may be obtained using data of just a single type. Secondly, the occurrence of these instances appears to be governed in large part by the specific type of integration used, the diagnostic comparison to which it is applied, and even differences in the

precise data used for a given modality. For example, although MKL was able to enhance discrimination of FEP from HCs by combining the fMRI contrast Suppression>Cross-Fixation during Suppression with sMRI and DTI data, it could not achieve the same accuracy using a different fMRI contrast (i.e. Initiation>Cross-Fixation during Initiation), suggesting that the specific contrast selected to classify subjects is an important factor.

5.4.1 Limitations

One of the primary limitations of the study stems from the fact that due to the relatively small number of different data types being integrated, the most computationally complex method, MKL could not in some ways, be thoroughly examined. However, taking as an example the 2000 plus single SVM classifier total, combined into a single MKL SVM used to enhance classification accuracy as reported by Damoulas et al., it is clear this is a figure highly unlikely to be matched with respect to any neuroimaging investigation of psychiatric illness, and certainly beyond the scope of this study. In this context it is worth noting that the data shown here represent a very real example of the neuroimaging information likely to be available to any given clinician, if not more. A second limitation is the fact that not all five data types reported in Chapter 4 could be integrated here, which would be of intuitive interest. As discussed in section 5.2.3, however, this decision was made in order to maximize the number of subjects available for integration thereby allowing more robust inferences. Third, since an integration of two data types using MV would result in a largely heuristic outcome (see section 5.2.2.4), any inference regarding classification enhancement could not be made entirely equally across all integrative approaches; though this was partially offset by the division between two, and three, data integration approaches. A fourth limitation of the study is

the fact that since the overall number of integrations tested was relatively small, it is not possible to draw any definitive conclusions regarding the potential ability of each integrative method to enhance classification accuracy, or not, as the case may be. Finally, since only relatively small subject numbers were used means that it is also not possible to draw a definite conclusion regarding the diagnostic potential of integrated SVM with respect to ARMS and FEP subjects for the data types combined here.

5.4.2 Conclusion

In the current study I have, for the first time, made a comparison between four distinct approaches in terms of their ability to enhance SVM classification accuracy of early-stage psychosis by integrating data from multiple neuroimaging modalities. Following individual application to three separate diagnostic comparisons related to the ARMS and FEP, it is evident that the specific integrative approach used, the number of data types integrated, and also the diagnostic comparison to which they are applied, all appear to have substantial impact on the integrated accuracy achieved. This information may potentially serve as preliminary foundation on which future clinical development of integrative methods for clinical purposes may be based. It remains however, that due to the substantial variability in the level of integrated results obtained, for the time being, it is the accuracies achieved using a single modality, single kernel classifier, that may be considered as a preferential first choice, both in terms of magnitude and ease of acquisition. To fully exploit the benefits of data integration therefore, further work using larger sample sizes is needed to identify the specific combination of data types and integrative approach, best suited to the clinical group(s) being examined.

Chapter 6

General Discussion

6.1 Summary of Main Findings

Studies of chronic schizophrenia (ChSz) patients have revealed alterations in neuroanatomy, neurofunction and neurocognition relative to HCs, which appear, at least in part, to be mediated genetically (Ellison-Wright & Bullmore, 2009; Honea, et al., 2005; Kindermann, et al., 1997; Walton, et al., 2012; Zakzanis et al., 2003). More recently, qualitatively similar, albeit less severe, changes have also been reported in the earliest stages of psychosis, namely, those who have experienced a recent first episode of psychosis (FEP) and those with an at-risk mental state (ARMS) (Brewer et al., 2006; Egerton, et al., 2011; Fusar-Poli, et al., 2011a; Fusar-Poli, et al., 2007b; Fusar-Poli, et al., 2012b; Mesholam-Gately, et al., 2009; Pettersson-Yeo, et al., 2011). To date however, the impact of such findings on clinical practice has been largely limited. This may at least in part be explained by the fact that the majority of previous studies used standard univariate analyses, which, allowing inferences at the group level only, are of limited use to clinicians who are required to make diagnoses at the level of the individual. Support vector machine (SVM) is one alternative analysis which, as a multivariate machine learning approach, i) allows inferences to be drawn at the level of the individual, and ii) provides sensitivity to spatially distributed, and subtle, effects in the input data. Together, these properties provide SVM with a high potential for translation to a clinical setting, an environment in which decisions must be made about individuals, based on a variety of information sources.

In this context, the overarching aim of the present thesis was to investigate FEP and ARMS subjects using a multi-modal approach, comprising biological and cognitive modalities previously shown to be altered in ChSz patients at group level. Critically, these modalities were examined using both a standard univariate, and also multivariate SVM, analysis. In this context, subject groups were first compared with respect to each data type using a standard univariate analysis, to test for the presence of any gross, focal, abnormalities that might exist between them. Next, the relative capacity of each data type to discriminate FEP and ARMS subjects from HCs, and each other, at the individual, rather than group, level was examined using the machine learning analysis, SVM. Crucially, by maintaining the same subject pairings for both the standard univariate and SVM multivariate analyses, it was also possible to draw an anecdotal comparison between the results of the two. Finally, based on evidence that SVM classification accuracy may be enhanced through the integration of data, thus enhancing its clinical potential, an empirical comparison of four different integrative methods was performed, and applied to different group comparisons. In this way, the aim was to observe whether an integrated accuracy could be obtained superior to that achieved by the single best modality, and explore the contributory factors that impact upon it.

As outlined in chapter 3, the results obtained through standard univariate analysis of the different data types proved relatively inconsistent both with similar studies of FEP and ARMS subjects (Egerton, et al., 2011; Fusar-Poli, et al., 2011a; Fusar-Poli, et al., 2007b), and also analogous studies involving patients with ChSz (Ellison-Wright & Bullmore, 2009; Honea, et al., 2005). Specifically, standard univariate analysis was unable to detect any significant difference in terms of either grey matter (GM) or neurofunction for any of the three diagnostic comparisons, nor in terms of white matter (WM) integrity with respect to FEP versus HC subjects, or FEP versus ARMS subjects.

Consistent with previous studies however, widespread reductions in WM integrity were detected in ARMS subjects relative to HCs (Carletti, et al., 2012). Furthermore, as hypothesised, standard analysis of the California verbal learning test results showed that FEP and ARMS subjects performed significantly worse than HCs with respect to multiple test domains, with a direct comparison between the two groups showing the performance of FEP and ARMS subjects to be effectively the same. However, as discussed, the absence of significant differences in terms of GM, WM or neurofunction, is not necessarily easily explained. Whilst a number of factors may have contributed (see section 3.4 for more detail), one possibility explored in particular was that differences were in fact present, but that, characterised by alterations of a subtle and diffuse nature, they were simply undetectable using a standard univariate analysis specifically intended to detect gross, focal abnormalities. To some extent, this notion was supported by the results of the subsequent multivariate SVM approach.

As described in Chapter 4, the results of a SVM analysis in conjunction with different individual data types, suggested that in terms of general discriminability FEP subjects were the most readily discriminable from HCs, followed by FEP from ARMS subjects, and finally ARMS from HCs, with significant (p<0.05) classification accuracies achievable using four, three and two of the five potential different data types for each comparison respectively. Specifically, these successful classifications were made between FEP and HC subjects using genotypic (67.86%), DTI (65.79%), fMRI (65.79% and 68.42%) and neuropsychological data (73.69%); between ARMS and HC subjects using sMRI (68.42%) and DTI data (65.79%); and, between FEP and ARMS subjects using sMRI (76.67%), fMRI (73.33%) and neuropsychological data (66.67%).

Based on an anecdotal comparison of the results from the two types of analysis, i.e. standard univariate and multivariate SVM, three trends may be observed. First, in a

number of instances the two analytical approaches appear to provide comparable results. For example, neither univariate nor multivariate analyses were able to find any distinct differences between FEP and HC subjects, FEP and ARMS subjects nor ARMS and HC subjects with respect to sMRI, DTI or fMRI data respectively. Conversely, both types of analysis were able to find features distinguishing ARMS and HC subjects, and FEP and ARMS subjects, based on DTI and neuropsychological data respectively. Second, it is possible that for specific cases a multivariate approach may indeed be able to detect differences, where a univariate analysis may not. In the current investigation for example, this was the case on five separate occasions, i.e. discriminating ARMS from HC subjects using sMRI data, FEP and HC subjects using DTI and fMRI data, and FEP from ARMS subjects using sMRI and fMRI data, together equalling the number of *concordant* results achieved by the two analyses (e.g. comparisons for which both analysis types could, or could not, detect a significant difference). Third, and in a reversal to trend two, for some specific cases, it may in fact be standard univariate analysis that is able to detect significant difference(s) in the data whilst a multivariate approach in comparison, may not. For example, whilst standard analysis was able to detect differences in the neuropsychological scores of ARMS subjects relative to HCs, in the current study, SVM could not.

Collectively, these findings are largely consistent with the notion that multivariate SVM analysis is an approach more sensitive to subtle and diffuse changes in the input data relative to standard statistical techniques (Brammer, 2009). However, contrary to this, the results also demonstrate that the relative sensitivities of the two may be frequently matched, in addition to occasions where standard techniques may be able to detect alteration(s), where SVM cannot. From a research specific perspective therefore, it seems advisable that rather than adopt an 'either/or' analytical approach, perhaps driven

by constraints of time and/or cost, where possible both analysis types should be used alongside one another, providing complementary information resulting from distinct statistical approaches, each particularly suited to detecting specific types of difference. In comparison, from a specifically clinical perspective, it should be reiterated that unlike standard univariate analyses, it is the multivariate properties inherent to SVM, allowing inferences to be drawn at the level of the individual, that lend it such high level potential for clinical translation. With this in mind, it is clearly evident that the successful classification accuracies reported in the present thesis, though significant, fall considerably short of the level that would be hoped for, and ultimately expected of, any real-world clinical use of SVM. Taking these modest results into account, some preliminary evidence is outlined in chapter 5, as to whether such issues may be surmountable based on the integration of different data types.

As detailed in the final experimental chapter, the results of an empirical comparison of four integrative methods, each used to derive a consensus SVM decision, suggested two important facts. First, whilst integration *per se* may have the potential to increase classification accuracy by up to 13% relative to the best single modality classification accuracy (BSMCA), in the current study such enhancement was observed to be the exception rather than the norm, with the level of integrated accuracy actually achieved varying substantially. Second, whether such enhancement was observed or not, seemed to be considerably influenced by a range of factors, including the number of data types being combined, the types of data being combined, in addition to the specific diagnostic comparison to which the classifier is applied. Taking these facts into consideration, it seems clear that whilst the high level of *potential* clinical translation remains for SVM, supported in principle by the theoretically considerable increase in accuracy shown here to be achievable through data integration, there remain a number of substantial

limitations that need to be addressed (the specifics of which are more fully addressed in section 6.5).

6.2 Study Strengths

Overall, the core strengths of this work are threefold: i) the use of a multi-modal approach comprising both biological and cognitive indices, ii) the application of both univariate and multivariate analyses, and iii) the joint examination of both ARMS and FEP subjects, representing those thought to be at risk of imminent psychosis, and recent post-psychosis onset respectively.

With respect to the first of these, preliminary evidence from a wide number of studies reporting alterations in ARMS and FEP subjects relative to HCs, across a range of dimensions, has suggested that both psychosis, and psychosis-risk, may be quantified using a variety of indices (see section 1.2). By employing a multimodal approach therefore, the results of the current thesis further contribute to the understanding of how these two clinically dissociable states manifest themselves, and, more importantly, provide some insight into how this manifestation is simultaneously reflected across different biological and cognitive indices, in the same individuals. For example, it is perhaps of interest to note, that whilst ARMS subjects differed significantly from HCs in terms of widespread reduction in WM integrity, no corresponding alteration was evident with respect to neurofunction. In addition to this, the fact data collection was performed on the same day, for the same subject, for every modality, meant that the issues related to testing subjects at different time points were also minimised. For example, since psychopathological score and neuroimaging data were collected on the same day, it follows that one may assume more readily that the two indices were

reflective of a common, underlying, overall mental state, than if the two measurements had been taken weeks, or months, apart.

With respect to the second and third strengths combined, the multivariate SVM analysis used here is an analytical approach that is becoming increasingly applied in the fields of psychiatric and neurological research, driven largely by the want of more clinically translatable results (Borgwardt & Fusar-Poli, 2012; Brammer, 2009; Matthews, et al., 2006; Orrù, et al., 2012). This is in comparison to the majority of studies conducted to date, which have predominantly taken a uni-modal approach, and typically contrasted either ARMS and HC subjects, or FEP and HC subjects. By applying this relatively novel analysis to multiple types of data, and with respect to both ARMS and FEP subjects at the same time therefore, the results generated provide an insight into i) the relative capacity of different types of data to discriminate ARMS and FEP subjects from HCs, and each other, and also ii) SVM's relative clinical potential with respect to diagnosis of early stage-psychosis, and the extent to which this specific analytical avenue should pursued over others.

In consideration of the second strength specifically furthermore, by employing both univariate and multivariate analyses alongside one another, the results provide a qualitative insight into the performance of each approach with respect to the same data set. As outlined above, this in turn may provide some evidence for fellow researchers of the relative benefits of using a SVM approach, or lack thereof, with the specific results found here perhaps reiterating that the more standard univariate analyses, which provide a similar, though subtly different, complementary perspective, should not be abandoned.

Similarly, with respect to the third strength specifically, the investigation of both ARMS and FEP subjects in the same study, using the same methodological and analytical

pipelines, has in itself a number of benefits. Firstly, from a general perspective, by studying those with a recent FEP, and those thought to be in its prodromal stage, i.e. the ARMS, many of the confounding factors associated with investigations of ChSz are minimised if not negated completely, including; effects of prolonged exposure to anti-psychotic medication, effects of institutionalisation and the effects of chronicity. Secondly, as per the motivations of the original investigations of ARMS and FEP subjects, by studying those thought to be in the earliest stages of psychosis it is hoped that the findings may, at some point in the future, ultimately contribute to the provision of earlier and more effective treatment intervention, which in turn may delay the onset and/or recurrence of frank psychosis, if not prevent it altogether. Thirdly, the current study format means that a more direct anecdotal comparison could be made between the relative differences observed between ARMS and HC subjects, and FEP and HC subjects, providing a potentially more accurate picture of the relationship between these two groups with respect to normal physiological function. In addition to this, the format also allowed the first direct cross sectional comparison between FEP and ARMS subjects using SVM, with the subsequent results representing a timely perspective relating to the discussion of whether the ARMS should be considered as a unique diagnostic category (Carpenter & van Os, 2011). For example, the finding that ARMS subjects differed from both HC and FEP subjects may potentially be used as added support to the notion of the ARMS as a distinct clinical entity.

6.3 Study Limitations

Overall, there were a number of limitations related to the investigation presented here which may be summarised by five main points; i) exposure to antipsychotic medication,

ii) sample size, iii) impossibility to use a univariate statistical test for genotype, iv) cross-sectional versus longitudinal, and v) single-, versus multi-, centre.

As previously discussed, the fact that the majority of FEP subjects recruited to the current study were not antipsychotic naïve perhaps represents its primary limitation. To date, it remains unclear as to the precise effects of antipsychotic medication on neuroanatomy and/or neurofunction (Fusar-Poli, et al., 2007a; Navari & Dazzan, 2009), and as such, the issue is complicated by the absence of any straightforward mechanism able to adjust for such confound; indeed, it is a problem applicable to the majority of FEP investigations, as well as psychiatric research more generally. Although correlation analyses found no evidence that medication was driving the successful discrimination of FEP subjects, from either ARMS or HC subjects (see section. 4.3.8), it remains possible that medication may have still contributed to the classifications in an unknown way.

A second limitation of the study is represented by the relatively small sample size used. Whilst a recent analysis suggested 16-35 to be the optimal group size for the purposes of classical inference (Friston, 2012), a more pressing issue in the current context was the reduction of subject pairs for analysis using SVM, when data of a particular type was unavailable for one of the subjects. As noted in Chapter 4, this had the biggest impact on the SVM analyses using genetic, and to a lesser extent, neuropsychological data. However, having said that, the fact remains that since *all* comparisons made here were performed using relatively modest sample sizes, it was, and is, not possible to draw a definite conclusion regarding the diagnostic potential with respect to ARMS and FEP subjects for the data types tested.

A third limitation of the study is represented by the inability to perform a univariate test for genotype, analogous to the multivariate analysis used, thus inhibiting an anecdotal

comparison of the results of the two analyses. This is due to the fact that in this case, the appropriate test would have been a genetic association test, examining the relative frequencies of the risk allele, of each SNP used here, in each group. Since such tests typically require hundreds, or thousands, of subjects to reliably detect a significant effect (Lewis & Knight, 2012), it was therefore not practically possible. It is worth noting however that in contrast to other modalities, the SNPs used here were selected specifically for the fact that they represented the summation of SNPs reported to date (as of June 14th 2011) as conferring increased risk for psychosis, thereby reducing the chance of any non-relevant SNPs been included. In addition to this, it may be further noted that the allelic frequency in the HC group for the SNPs used were all in Hardy Weinberg Equilibrium, a standard quality control step in molecular genetics, providing confirmation that genetically speaking the case-controls used here could be considered representative of the general population (Lunetta, 2008).

The fourth issue is defined by the limitations associated with performing a cross-sectional, rather than longitudinal, study. First, and most presciently, this meant that no predictive inference could be made for the ARMS subjects involved with respect to psychosis transition. Secondly, since the investigation of FEP and ARMS subjects is ultimately driven by the objective of delaying, and/or, preventing the onset of ChSz, ideally one would like to observe how the measures of the different biological and cognitive indices change with respect to time and associated psychopathology. In this way, it would be possible to obtain a more precise understanding of the exact alterations associated specifically with the ARMS, with FEP, and with the transition from the former to the latter, based on the absence of confounding factors arising from the use of different subjects. However, having said that, by including ARMS, FEP and HC subjects, the format of the current study provided the next best alternative to a

longitudinal study within the available time and funding constraints, enabling similar questions to be asked albeit less directly. In addition to this, one practical benefit of the cross sectional nature of the current study meant that it was not dependent on the ARMS subjects involved in the study transitioning to a FEP, which, based on a recently reported transition rate of 36% within 3 years (Fusar-Poli, et al., 2012a), is by no means guaranteed.

A fifth limitation of the study was the fact that it was based on data collected from a single centre, rather than from multiple centres. As with any study restricted to a single location, this meant that any inference made regarding the applicability of the findings to ARMS and FEP subjects in general, was inherently less strong. Furthermore, this also meant that any successful SVM decision function differentiating ARMS and FEP subjects from HCs, or each other, derived through leave-one-out cross validation, was tested with respect to the one data set only. Whilst this has been shown to provide a relatively robust, and unbiased, measure of generalizability estimation (Lemm, et al., 2011), it would be of great interest to measure the performance of the classifier with respect to a different data set acquired elsewhere. This would be of particular value from the perspective of the potential clinical translation of SVM, for which in a real-world scenario subject measurements are most likely to have been obtained from a variety of locations. The fact data was acquired in the current study using the same equipment, at the same site, for all subjects however, did minimise the potential confounding impact associated with data collected from multiple sites, by multiple people, using differing hardware and/or software.

6.4 Relationship to Previous Work

Outlined in Chapter 1, previous studies of FEP and ARMS subjects have revealed alterations in a host of biological and cognitive indices qualitatively similar, albeit less severe, to those observed in ChSz. Mainly employing univariate analyses, each individual study has typically reported gross abnormalities, localised to a specific cognitive subcomponent or neuroanatomical region, in each group relative to HCs. Since the precise locality, and/or direction, of these differences has not always been constant between studies however (Egerton, et al., 2011; Fusar-Poli, et al., 2011a; Hawkins et al., 2008; Mesholam-Gately, et al., 2009; Pettersson-Yeo, et al., 2011), one possible inference is that FEP and the ARMS are associated with diffuse, rather than solely localised, changes. To some extent, this notion is supported by the significant SVM classifications found here, successfully discriminating FEP and ARMS from HCs, and each other. For example, whilst gross focal abnormalities in WM integrity were evident between ARMS and HC subjects using univariate analyses, significant results found using multivariate SVM analysis of the same data suggest that such changes may be only one feature of a more diffuse pattern of alteration. Furthermore, the fact SVM was able to discriminate individuals between groups using some particular data types whilst univariate analysis could not, suggests that these diffuse alterations may be characterised by subtle changes, in addition to those gross differences previously reported. Taken together therefore, the results here are consistent with the notion inferred by previous studies collectively that alterations in early stage psychosis are characterised by widespread differences, of a potentially subtle nature, in addition to gross localised abnormalities.

As highlighted above however, the magnitude of the diagnostic classification accuracies obtained here, though encouraging, are at best modest. Generally speaking, this is relatively inconsistent with the few previous studies to have examined subjects with an ARMS or FEP using SVM (Koutsouleris, et al., 2011b; Koutsouleris, et al., 2009a; Sun, et al., 2009). As discussed in section 4.4, this could due to a number of possible methodological differences. Selecting two areas of potential difference from those detailed in section 4.4 for example, such areas could include i) differences in subject assessment criteria, and/or ii) differences in SVM implementation. To elaborate on these more fully, with respect to the first, as mentioned in section 1.1.2, there currently exist a number of assessment tools intended to categorise those with an ARMS. Whilst preliminary studies suggest a substantial rate of concordance between them (Chuma & Mahadun, 2011), it remains possible that differences in assessment criteria may have resulted in the inconsistent findings. In the studies by Koutsouleris et al. for example the ARMS subjects were sub-categorised into those believed to be at imminent risk of psychosis transition, and those at less imminent risk. This was based however on a combination of different assessment tools, comprising an index of basic symptoms taken from the Bonn Scale for Assessment of Basic Symptoms (BSABS), and definitions of attenuated psychotic and brief limited intermittent psychotic symptoms as defined by the Personal Assessment and Crisis Evaluation (PACE) criteria. In comparison, the ARMS subjects recruited here were categorized by clinical psychiatrists using the PACE criteria only, in accordance with the comprehensive assessment of at-risk mental states (CAARMS), with an ARMS categorization inherently defining that individual as at risk of imminent psychosis. It is possible therefore, that had there been greater consistency in the clinical assessment tools used, more consistent results may have been obtained.

With respect to the second possibility, the progressive and increasingly widespread use of multivariate analyses in the field of psychiatry is a relatively new development, and as such, there is currently no established methodology agreed to be the best. Hence, although a different study may also be reported as using a 'SVM analysis', the precise method by which the algorithm's parameters were selected, and the final decision function implemented could vary greatly. These could include for example the manner in which features are constructed for input into SVM (i.e. the method by which the data was preprocessed and the format of the subsequent output data used as input for SVM), the settings chosen for the variable parameters, such as kernel type, the level of the parameter $\boldsymbol{C}$, or the regularization norm used for the slack variables ($\boldsymbol{\xi}$), and/or the use, or lack thereof, of a feature selection step (Schölkopf & Smola, 2002). Collectively therefore, each of these possibilities represents a simple demonstration of the manner in which inconsistencies may arise between two 'SVM' studies, that may in turn have resulted in the differences in results.

With respect to classifier integration specifically, the fact the data showed that classification accuracy may be enhanced using this approach is consistent with the one previous study investigating ChSz, in addition to the analogous studies investigating Alzheimer's disease (Fan, et al., 2008; Hinrichs, et al., 2011; Yang, et al., 2010; Zhang, et al., 2011). The current data expands upon this however, by demonstrating that whilst substantial increase is possible, such instances appear to represent the exception rather than the norm, with the levels of integrated accuracy achieved varying substantially. Furthermore, it confirms that the ultimate level of classification accuracy attainable with respect to psychosis may be influenced by a range of factors (outlined above).

6.5 SVM: A Potential Real-World Clinical Tool?

In spite of these methodological issues, it remains that the significant classification accuracies found here and in past studies, albeit less than 100%, remain consistent with the notion that SVM, *per se*, has a high level of clinical translation potential able to inform assessment of those thought to have an ARMS or FEP. However, within this context, it should be acknowledged that the 'diagnostic' SVM format, as used in the current study, has a number of limitations.

First, the very fact that neither ARMS, nor FEP, subjects have yet been successfully discriminated from HCs with 100% accuracy – by this, or any other study – raises the issue that any translational implementation of SVM must account for the cost of erroneous classification. For example, the cost of misclassifying someone unwell as healthy, may have a greater detrimental impact than if someone healthy was classified as unwell. As such, a classifier with excellent sensitivity, but only good specificity, may be preferred to one with excellent specificity but only good sensitivity.

Second, it should be noted that the application of supervised SVM, as used here, may only reach the same level of diagnostic accuracy as traditional, interview based, methods of clinical assessment. This is due to the fact that the initial development of the SVM decision function is derived from the distinction between subjects in the training data whose categories are pre-defined by the researcher, which in turn are based on traditional clinical assessment. From this diagnostic perspective therefore, it would hence be an expression of logical confusion to expect an SVM algorithm to allow better diagnostic classification than traditional clinical assessment from which it was developed.

Third, whilst inferences drawn from SVM may be made at the level of the individual, and hence be potentially clinically useful, it is also true that clinicians are often required to make prompt clinical decisions which may be inconsistent with the timeframe needed to collect and prepare the data to which SVM is applied, leading to a potentially impractical and harmful treatment delay for the patient.

Fourth, the application of SVM comprises a series of steps requiring both technical and computational expertise beyond the capabilities of most clinical units, creating an inherent restriction on its use governed by a clinic's financial capacity.

Finally, it should be noted that SVM would not be suitable for those subjects identified as having gross abnormalities that are comorbid to their psychiatric, or neurological, illness. This is because by effectively introducing noise into the data the presence of such an abnormality would have the potential to malignly impact the accuracy of a classifier if it is being developed for the first time, or lead to a false classification if using a decision function that has already been generated.

Nevertheless, taking these limitations into account, the diagnostic format of SVM in conjunction with multi-modal data, of which the results presented here provide a partial proof of concept, may still have its uses. For example, a dedicated algorithm could be generated representing a 'gold-standard', based on the opinions of internationally recognized experts in the field. This algorithm could then be employed to help inform diagnosis of subjects in clinics around the world, where such high levels of psychiatric expertise may not be available. In addition to this, a second more diverse usage may be found in a forensic setting, where supervised SVM could be utilized in order to reduce controversy in evaluations of mental insanity and to minimize errors in the detection of malingering (Sartori, et al., 2011).

However, in the context of the multiple limitations listed above, it may be reasonably argued that the primary strength of SVM lies not as a diagnostic aid, but rather, as a predictive tool applicable to disease prognosis and/or treatment outcome. In the field of psychosis for example, preliminary data from longitudinal studies have suggested that it is possible to retrospectively predict, with a significant degree of accuracy, ARMS subjects who will, and will not, transition to a FEP (Koutsouleris et al., 2011a) based on alterations in neuroanatomy. More generally, this application of SVM has also been strongly supported by analogous studies investigating the hypothesized prodrome of probable dementia of Alzheimer's type (PDAT), mild cognitive impairment (MCI). Specifically, classifiers able to successfully predict (retrospectively) MCI subjects who would, and would not, develop frank PDAT, were generated using data from the Alzheimer's disease neuroimaging initiative (ADNI) database (Chincarini et al., 2011; Zhang, et al., 2011). Critically, the ADNI database comprises a collection of data acquired using different scanners across multiple sites, therefore providing strong evidence that at least in principle, a SVM classifier generated in one centre can indeed be generalized for use on data acquired at a different centre.

With respect to the prediction of treatment outcome comparatively, similarly encouraging results have also been reported. A study investigating response to Clozapine treatment in ChSz patients for example found that responders could be discriminated from non-responders based on pre-treatment fluctuations in neuronal activity, measured using electroencephalography, with up to 85% accuracy (Khodayari-Rostamabad et al., 2010). Furthermore, such results have also been reported with respect to other psychiatric illness', including major depression, with treatment response to both pharmacological, and non-pharmacological interventions, accurately predicted using sMRI and fMRI respectively (Costafreda et al., 2009b; Gong, et al., 2011).

In consideration of these results with respect to psychosis specifically, two areas emerge in which SVM could be of real use. First, given the ability to predict with relative confidence those most at risk of transitioning to FEP, treatment intervention could be applied earlier, and in a more targeted way, maximizing efficacy and minimizing unnecessary treatment for those not in need of it. Second, given the ability to predict who will, and will not, respond well to a given treatment opens the door to personalized medicine, again, maximizing efficacy, and minimizing the distressing effects of refractory psychosis and unpleasant side effects occurring in the absence of any clinical benefit.

When considering the future use of SVM for clinical purposes, it is also perhaps worth mentioning alternative forms of machine learning which have yet to be applied in the psychiatric field. Whilst a full discussion exploring the various types is beyond the scope this thesis, two alternative approaches of emerging promise include probabilistic machine learning, and support vector regression. With respect to the first, rather than providing a categorical classification for each subject as per the SVM used here, probabilistic machine learning returns an estimate of the probability that a given subject belongs to each category (e.g. 80% probability this subject is a HC, 20% probability they are a patient), thereby quantifying the uncertainty of each prediction. Of this analytical type specifically, two methods include Gaussian processes (Rasmussen & Williams, 2006) and relevance vector machines (Tipping, 2001), with preliminary data from non-psychiatric neuroimaging studies, involving both neuroanatomical and neurofunctional data, providing encouraging results (Marquand et al., 2010; Phillips et al., 2011). With respect to the second main type of analysis in comparison, machine learning regression aims to predict a continuous outcome variable, such as symptom severity, rather than a categorical class label as obtained with SVM. Like probabilistic

machine learning, recent evidence from non-psychiatric studies, have provided encouraging results in the context of both normal health (Franke et al., 2010), and also neurological disease (Stonnington et al., 2010). Based on these encouraging preliminary results therefore, both probabilistic and regression machine learning methods appear to have high potential for translation into a clinical setting, both psychiatric and general, in which physicians are required to make a balanced treatment decision based on an objective consideration of the potential risks and benefits for the patient.

6.6 Future Work

Based on the findings presented here, and in light of the limitations both of the study, and SVM methodology more generally, there are a number of interesting avenues arising from this work that could be investigated by future studies.

It would be of interest for example, to investigate how well the successful classifiers obtained here generalized to data sets obtained from other research sites. This would in turn provide a greater estimate of the true accuracy of the successful classifiers generated, and thus, a more sensitive indicator of SVM's *diagnostic* potential based on the data types investigated here.

In the longer term, I plan to continue follow up of the ARMS subjects who participated in the study in order to identify any of those who transition to psychosis. In particular, I plan to combine the present data set with other longitudinal data sets acquired at the Institute of Psychiatry in order to be able to compare those ARMS subjects go on to develop psychosis against those who remain stable with sufficient statistical power. In addition to this, it would also be of interest to collaborate with other sites conducting similar follow up studies, in order to create a large pool of ARMS subjects, for which a

comprehensive SVM analysis, based on an even larger number of subjects collated from multiple research sites, could be made, similar to the recent univariate analysis conducted by Mechelli et al. (2011). This would in turn provide greater insight into the true *predictive* potential for SVM, and a more realistic estimate of its potential for clinical translation with respect to prognosis. It is perhaps of interest, that as of 18th of September 2012, according to their medical records, none of the ARMS subjects recruited in this study had developed frank psychosis.

A third avenue arising from the work here, would involve a similar study setup, but instead using alternative types of input data on the basis that they may index different components of the same underlying mental state. For example, such data types could include other neuroimaging modalities, such as positron emission tomography, or perfusion data, or, cerebro-spinal fluid markers.

With respect to the field more generally, there are also a number of potential research avenues that arise from this work. It would be of interest for example to investigate whether the same methods used here, could also be used to differentiate between different psychiatric and/or neurological diseases, which may share similar, or comorbid, symptom traits, providing further support for, or against, the diagnostic use of SVM. Another avenue related to ARMS and FEP research specifically, stems from the difficulty of comparing the results of studies which differed with respect to subject assessment. One obvious solution to this is the adoption of a more consistent, cohesive approach by the different research centres, with respect to the assessment protocols used. As discussed, whilst preliminary evidence suggests a high level of concordance between the different measures used to assess someone as having an ARMS (Chuma & Mahadun, 2011), it remains that the assessments are not in fact identical. By adopting a more standardized approach therefore, the ability to compare results from different research

groups would be greatly strengthened. In addition, extending this idea one step further, if consistency was also found between the different psychometric, and psychopathological, scales used by each group, this too would greatly facilitate the pooling of data from different sites. Lastly, since SVM in fact represents only one branch of a wider field of supervised learning algorithms, another avenue for future work is the continued application of classifiers derived from alternative branches, including for example artificial neural network, decision tree and random forest approaches (Caruana & Niculescu-Mizil, 2006), in order to help clarify the optimal classification method with respect to early stage-psychosis.

6.7 Conclusion

To conclude, the results presented in this thesis are in part consistent with the notion that, relative to HCs, ARMS and FEP subjects are associated with genetic, neuroanatomical, neurofunctional and cognitive alterations that are qualitatively similar to, albeit less severe than, those previously observed in ChSz subjects. With specific respect to neuroanatomy and neurofunction moreover, they suggest such alterations may be both subtle, and spatially diffuse, and in some cases, may even be detectable in the absence of gross, focal abnormalities. Whilst the modest classification accuracies observed here suggest that the different modalities investigated have only limited diagnostic power with respect to early-stage psychosis, it remains that they may be able to provide useful information for predicting conversion to psychosis or treatment outcome, a prospect which could be investigated in future studies.

REFERENCES

Addington, J., Chaves, A., & Addington, D. (2006). Diagnostic stability over one year in first-episode psychosis. *Schizophr Res, 86*(1-3), 71-75.

Allen, P., Chaddock, C. A., Howes, O. D., Egerton, A., Seal, M. L., Fusar-Poli, P., et al. (2011). Abnormal Relationship Between Medial Temporal Lobe and Subcortical Dopamine Function in People With an Ultra High Risk for Psychosis. *Schizophrenia bulletin*.

Allen, P., Mechelli, A., Stephan, K. E., Day, F., Dalton, J., Williams, S., et al. (2008). Fronto-temporal interactions during overt verbal initiation and suppression. *Journal of cognitive neuroscience, 20*(9), 1656-1669.

Allen, P., Stephan, K. E., Mechelli, A., Day, F., Ward, N., Dalton, J., et al. (2010). Cingulate activity and fronto-temporal connectivity in people with prodromal signs of psychosis. *NeuroImage, 49*(1), 947-955.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev. ed.). Washington, DC: American Psychiatric Association.

Ananth, H., Popescu, I., Critchley, H. D., Good, C. D., Frackowiak, R. S. J., & Dolan, R. J. (2002). Cortical and subcortical gray matter abnormalities in schizophrenia determined through structural magnetic resonance imaging with optimized volumetric voxel-based morphometry. *The American journal of psychiatry, 159*(9), 1497-1505.

Andersson, J. L. R., Jenkinson, M., & Smith, S. (2007a). Non-linear optimisation. FMRIB technical report TR07ja1 from http://www.fmrib.ox.ac.uk/analysis/techrep.

Andersson, J. L. R., Jenkinson, M., & Smith, S. (2007b). Non-linear registration, aka Spatial normalisation. FMRIB technical report TR07JA2 from http://www.fmrib.ox.ac.uk/analysis/techrep.

Andreasen, N. C., Ehrhardt, J. C., Swayze, V. W., Alliger, R. J., Yuh, W. T., Cohen, G., et al. (1990). Magnetic resonance imaging of the brain in schizophrenia. The pathophysiologic significance of structural abnormalities. *Arch Gen Psychiatry, 47*(1), 35-44.

Antonova, E., Kumari, V., Morris, R., Halari, R., Anilkumar, A., Mehrotra, R., et al. (2005). The relationship of structural alterations to cognitive deficits in schizophrenia: a voxel-based morphometry study. *Biological psychiatry, 58*(6), 457-467.

Arcuri, S. M., Rabe-Hesketh, S., Morris, R. G., & McGuire, P. K. (2001). Regional variation of cloze probabilities for sentence contexts. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc, 33*(1), 80-90.

Arimura, H., Yoshiura, T., Kumazawa, S., Tanaka, K., Koga, H., Mihara, F., et al. (2008). Automated method for identification of patients with Alzheimer's disease based on three-dimensional MR images. *Acad Radiol, 15*(3), 274-284.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage, 38*(1), 95-113.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry--the methods. *Neuroimage, 11*(6 Pt 1), 805-821.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage, 26*(3), 839-851.

Basser, P. J., & Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J Magn Reson B, 111*(3), 209-219.

Benetti, S., Mechelli, A., Picchioni, M., Broome, M., Williams, S., & McGuire, P. (2009). Functional integration between the posterior hippocampus and prefrontal cortex is impaired in both first episode schizophrenia and the at risk mental state. *Brain, 132*(Pt 9), 2426-2436.

Bilder, R. M., Goldman, R. S., Robinson, D., Reiter, G., Bell, L., Bates, J. A., et al. (2000). Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates. *Am J Psychiatry, 157*(4), 549-559.

Bloemen, O. J. N., de Koning, M. B., Schmitz, N., Nieman, D. H., Becker, H. E., de Haan, L., et al. (2009). White-matter markers for psychosis in a prospective ultra-high-risk cohort. *Psychological medicine*, 1-8.

Bloom, P. A., Fishchler, I. (1980). Completion norms for 329 sentence contexts. *Memory and Cognition* 631-642.

Bonilha, L., Molnar, C., Horner, M. D., Anderson, B., Forster, L., George, M. S., et al. (2008). Neurocognitive deficits and prefrontal cortical atrophy in patients with schizophrenia. *Schizophrenia research, 101*(1-3), 142-151.

Borgwardt, S., & Fusar-Poli, P. (2012). Third-generation neuroimaging in early schizophrenia: translating research evidence into clinical utility. *Br J Psychiatry, 200*, 270-272.

Borgwardt, S. J., McGuire, P. K., Aston, J., Berger, G., Dazzan, P., Gschwandtner, U., et al. (2007a). Structural brain abnormalities in individuals with an at-risk mental state who later develop psychosis. *Br J Psychiatry Suppl, 51*, s69-75.

Borgwardt, S. J., McGuire, P. K., Aston, J., Gschwandtner, U., Pflüger, M. O., Stieglitz, R.-D., et al. (2008). Reductions in frontal, temporal and parietal volume associated with the onset of psychosis. *Schizophrenia research, 106*(2-3), 108-114.

Borgwardt, S. J., Picchioni, M. M., Ettinger, U., Toulopoulou, T., Murray, R., & McGuire, P. K. (2010). Regional gray matter volume in monozygotic twins concordant and discordant for schizophrenia. *Biological psychiatry, 67*(10), 956-964.

Borgwardt, S. J., Riecher-Rössler, A., Dazzan, P., Chitnis, X., Aston, J., Drewe, M., et al. (2007b). Regional gray matter volume abnormalities in the at risk mental state. *Biological psychiatry, 61*(10), 1148-1156.

Brammer, M. (2009). The role of neuroimaging in diagnosis and personalized medicine--current position and likely future directions. *Dialogues Clin Neurosci, 11*(4), 389-396.

Brewer, W. J., Francey, S. M., Wood, S. J., Jackson, H. J., Pantelis, C., Phillips, L. J., et al. (2005). Memory impairments identified in people at ultra-high risk for psychosis who later develop first-episode psychosis. *The American journal of psychiatry, 162*(1), 71-78.

Brewer, W. J., Wood, S. J., Phillips, L. J., Francey, S. M., Pantelis, C., Yung, A. R., et al. (2006). Generalized and specific cognitive performance in clinical high-risk cohorts: a review highlighting potential vulnerability markers for psychosis. *Schizophrenia bulletin, 32*(3), 538-555.

Broome, M. R., Matthiasson, P., Fusar-Poli, P., Woolley, J. B., Johns, L. C., Tabraham, P., et al. (2009). Neural correlates of executive function and working memory in the 'at-risk mental state'. *Br J Psychiatry, 194*(1), 25-33.

Broome, M. R., Woolley, J. B., Johns, L. C., Valmaggia, L. R., Tabraham, P., Gafoor, R., et al. (2005). Outreach and support in south London (OASIS): implementation of a clinical service for prodromal psychosis and the at risk mental state. *Eur Psychiatry, 20*(5-6), 372-378.

Brown, G. G., Lee, J.-S., Strigo, I. A., Caligiuri, M. P., Meloy, M. J., & Lohr, J. (2011). Voxel-based morphometry of patients with schizophrenia or bipolar I disorder: a matched control study. *Psychiatry research, 194*(2), 149-156.

Burdick, K. E., Kamiya, A., Hodgkinson, C. A., Lencz, T., DeRosse, P., Ishizuka, K., et al. (2008). Elucidating the relationship between DISC1, NDEL1 and NDE1 and the risk for schizophrenia: evidence of epistasis and competitive binding. *Hum Mol Genet, 17*(16), 2462-2473.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*.

Burgess, P. W., Shallice, T. (1996). Response suppression and initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, 263-272.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med, 39*(6), 855-864.

Carletti, F., Woolley, J. B., Bhattacharyya, S., Perez-Iglesias, R., Fusar Poli, P., Valmaggia, L., et al. (2012). Alterations in White Matter Evident Before the Onset of Psychosis. *Schizophrenia bulletin*.

Carpenter, W. T., & van Os, J. (2011). Should attenuated psychosis syndrome be a DSM-5 diagnosis? *The American journal of psychiatry, 168*(5), 460-463.

Caruana, R., & Niculescu-Mizil, N. (2006). An Empirical Comparison of Supervised Learning Algorithms Using Different Performance Metrics, *Proc. 23rd Int'l Conf. Machine Learning (ICML '06)* (pp. 161-168).

Catani, M., Craig, M. C., Forkel, S. J., Kanaan, R., Picchioni, M., Toulopoulou, T., et al. (2011). Altered integrity of perisylvian language pathways in

schizophrenia: relationship to auditory hallucinations. *Biological psychiatry, 70*(12), 1143-1150.

Chan, R. C. K., Chen, E. Y. H., Cheung, E. F. C., Chen, R. Y. L., & Cheung, H. K. (2006a). The components of executive functioning in a cohort of patients with chronic schizophrenia: a multiple single-case study design. *Schizophrenia research, 81*(2-3), 173-189.

Chan, R. C. K., Chen, E. Y. H., Cheung, E. F. C., & Cheung, H. K. (2004). Executive dysfunctions in schizophrenia. Relationships to clinical manifestation. *Eur Arch Psychiatry Clin Neurosci, 254*(4), 256-262.

Chan, R. C. K., Chen, E. Y. H., & Law, C. W. (2006b). Specific executive dysfunction in patients with first-episode medication-naïve schizophrenia. *Schizophrenia research, 82*(1), 51-64.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol., 2*(3), 1-27.

Chang, L.-C., Jones, D. K., & Pierpaoli, C. (2005). RESTORE: robust estimation of tensors by outlier rejection. *Magn Reson Med, 53*(5), 1088-1095.

Chen, X., Lee, G., Maher, B. S., Fanous, A. H., Chen, J., Zhao, Z., et al. (2011). GWA study data mining and independent replication identify cardiomyopathy-associated 5 (CMYA5) as a risk gene for schizophrenia. *Mol Psychiatry, 16*(11), 1117-1129.

Cheung, V., Cheung, C., McAlonan, G. M., Deng, Y., Wong, J. G., Yip, L., et al. (2008). A diffusion tensor imaging study of structural dysconnectivity in never-medicated, first-episode schizophrenia. *Psychol Med, 38*(6), 877-885.

Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., et al. (2011). Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *Neuroimage, 58*(2), 469-480.

Chua, S. E., Cheung, C., Cheung, V., Tsang, J. T. K., Chen, E. Y. H., Wong, J. C. H., et al. (2007). Cerebral grey, white matter and csf in never-medicated, first-episode schizophrenia. *Schizophrenia research, 89*(1-3), 12-21.

Chuma, J., & Mahadun, P. (2011). Predicting the development of schizophrenia in high-risk populations: systematic review of the predictive validity of prodromal criteria. *Br J Psychiatry, 199*(5), 361-366.

Cichon, S., Mühleisen, T. W., Degenhardt, F. A., Mattheisen, M., Miró, X., Strohmaier, J., et al. (2011). Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet, 88*(3), 372-381.

Cocchi, L., Walterfang, M., Testa, R., Wood, S. J., Seal, M. L., Suckling, J., et al. (2009). Grey and white matter abnormalities are associated with impaired spatial working memory ability in first-episode schizophrenia. *Schizophrenia research, 115*(2-3), 163-172.

Collins, A. L., Kim, Y., Sklar, P., Consortium, I. S., O'Donovan, M. C., & Sullivan, P. F. (2012). Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychol Med, 42*(3), 607-616.

Collette, F., Van der Linden, M., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2001). The functional anatomy of inhibition processes investigated with the Hayling task. *NeuroImage, 14*(2), 258-267.

Cooke, M. A., Fannon, D., Kuipers, E., Peters, E., Williams, S. C., & Kumari, V. (2008). Neurological basis of poor insight in psychosis: a voxel-based MRI study. *Schizophrenia research, 103*(1-3), 40-51.

Coren, S., Porac, C., & Duncan, P. (1979). A behaviorally validated self-report inventory to assess four types of lateral preference (Vol. 1:1, pp. 55-64): Journal of Clinical Neuropsychology.

Cortes, C., & Vapnik, V. (1995). Support-vector networks (Vol. 20, pp. 273-297): Machine Learning.

Costafreda, S. G., Chu, C., Ashburner, J., & Fu, C. H. Y. (2009a). Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS ONE, 4*(7), e6353.

Costafreda, S. G., Fu, C. H. Y., Picchioni, M., Toulopoulou, T., McDonald, C., Kravariti, E., et al. (2011). Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry, 11*, 18.

Costafreda, S. G., Khanna, A., Mourao-Miranda, J., & Fu, C. H. Y. (2009b). Neural correlates of sad faces predict clinical remission to cognitive behavioural therapy in depression. *Neuroreport, 20*(7), 637-641.

Crossley, N. A., Mechelli, A., Fusar-Poli, P., Broome, M. R., Matthiasson, P., Johns, L. C., et al. (2009). Superior temporal lobe dysfunction and frontotemporal dysconnectivity in subjects at risk of psychosis and in first-episode psychosis. *Hum Brain Mapp, 30*(12), 4129-4137.

Damoulas, T., & Girolami, M. A. (2008). Probabilistic multi-class multi-kernel learning: on protein fold recognition and remote homology detection. *Bioinformatics, 24*(10), 1264-1270.

Davatzikos, C., Fan, Y., Wu, X., Shen, D., & Resnick, S. M. (2008). Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging, 29*(4), 514-523.

Davatzikos, C., Shen, D., Gur, R. C., Wu, X., Liu, D., Fan, Y., et al. (2005). Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. *Arch Gen Psychiatry, 62*(11), 1218-1227.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). California Verbal Learning Test: Adult version. Manual, *The Psychological Corporation*. San Antonio, TX.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (2000). California Verbal Learning Test, second Edition. Adult Version, *The Psychological Corporation*. San Antonio, TX.

Delis, D. C., & Fridlund, A. J. (2000). CVLT-II Comprehensive Scoring System and Computerized Report. San Antonio, TX: The Psychological Corporation.

Di, X., Chan, R. C. K., & Gong, Q.-Y. (2009). White matter reduction in patients with schizophrenia as revealed by voxel-based morphometry: an activation likelihood estimation meta-analysis. *Prog Neuropsychopharmacol Biol Psychiatry, 33*(8), 1390-1394.

Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G. B., & Collins, D. L. (2008). MRI-based automated computer classification of probable AD versus normal controls. *IEEE Trans Med Imaging, 27*(4), 509-520.

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., et al. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci USA, 98*(12), 6917-6922.

Egerton, A., Borgwardt, S., Tognin, S., Howes, O., Mc Guire, P., & Allen, P. (2011). An overview of functional, structural and neurochemical imaging studies in individuals with a clinical high risk of psychosis, *Neuropsychiatry* (pp. 477-493).

Ellison-Wright, I., & Bullmore, E. (2009). Meta-analysis of diffusion tensor imaging studies in schizophrenia. *Schizophrenia research, 108*(1-3), 3-10.

Ellison-Wright, I., Glahn, D. C., Laird, A. R., Thelen, S. M., & Bullmore, E. (2008). The anatomy of first-episode and chronic schizophrenia: an anatomical likelihood estimation meta-analysis. *Am J Psychiatry, 165*(8), 1015-1023.

Esslinger, C., Walter, H., Kirsch, P., Erk, S., Schnell, K., Arnold, C., et al. (2009). Neural mechanisms of a genome-wide supported psychosis variant. *Science, 324*(5927), 605.

Fan, Y., Resnick, S. M., Wu, X., & Davatzikos, C. (2008). Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage, 41*(2), 277-285.

Ferreira, M. A. R., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., et al. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet, 40*(9), 1056-1058.

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M. E., & Clare, L. (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychol Rev, 15*(2), 73-95.

Frackowiak, R. S., Friston, K., Frith, C., Dolan, R. J., Price, C., Zeki, S., et al. (2004). *Human Brain Function* (Second Edition ed.): Elsevier.

Franke, K., Ziegler, G., Klöppel, S., Gaser, C., & Initiative, A. s. D. N. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *NeuroImage, 50*(3), 883-892.

Friedman, J. I., Tang, C., Carpenter, D., Buchsbaum, M., Schmeidler, J., Flanagan, L., et al. (2008). Diffusion tensor imaging findings in first-episode and chronic schizophrenia patients. *The American journal of psychiatry, 165*(8), 1024-1032.

Friis, S., Sundet, K., Rund, B. R., Vaglum, P., & McGlashan, T. H. (2002). Neurocognitive dimensions characterising patients with first-episode psychosis. *The British journal of psychiatry Supplement, 43*, s85-90.

Friston, K. (2012). Ten ironic rules for non-statistical reviewers. *NeuroImage, 61*(4), 1300-1310.

Friston, K., Ashburner, J., & Frith, C. (1995). Spatial registration and normalization of images, *Human Brain Mapping*.

Fu, C. H. Y., Mourao-Miranda, J., Costafreda, S. G., Khanna, A., Marquand, A. F., Williams, S. C. R., et al. (2008). Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. *Biol Psychiatry, 63*(7), 656-662.

Fusar-Poli, P., Bonoldi, I., Yung, A. R., Borgwardt, S., Kempton, M. J., Valmaggia, L., et al. (2012a). Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk. *Arch Gen Psychiatry, 69*(3), 220-229.

Fusar-Poli, P., Borgwardt, S., Crescini, A., Deste, G., Kempton, M. J., Lawrie, S., et al. (2011a). Neuroanatomy of vulnerability to psychosis: a voxel-based meta-analysis. *Neuroscience and biobehavioral reviews, 35*(5), 1175-1185.

Fusar-Poli, P., Broome, M. R., Matthiasson, P., Williams, S. C. R., Brammer, M., & McGuire, P. K. (2007a). Effects of acute antipsychotic treatment on brain activation in first episode psychosis: an fMRI study. *Eur Neuropsychopharmacol, 17*(6-7), 492-500.

Fusar-Poli, P., Crossley, N., Woolley, J., Carletti, F., Perez-Iglesias, R., Broome, M., et al. (2011b). Gray matter alterations related to P300 abnormalities in subjects at high risk for psychosis: longitudinal MRI-EEG study. *NeuroImage, 55*(1), 320-328.

Fusar-Poli, P., Perez, J., Broome, M., Borgwardt, S., Placentino, A., Caverzasi, E., et al. (2007b). Neurofunctional correlates of vulnerability to psychosis: a systematic review and meta-analysis. *Neurosci Biobehav Rev, 31*(4), 465-484.

Fusar-Poli, P., Smieskova, R., Serafini, G., Politi, P., & Borgwardt, S. (2012b). Neuroanatomical markers of genetic liability to psychosis and first episode psychosis: A voxelwise meta-analytical comparison. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry*.

Gasparotti, R., Valsecchi, P., Carletti, F., Galluzzo, A., Liserre, R., Cesana, B., et al. (2009). Reduced fractional anisotropy of corpus callosum in first-contact, antipsychotic drug-naive patients with schizophrenia. *Schizophrenia research, 108*(1-3), 41-48.

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.-S., et al. (2009). Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage, 47*(4), 1476-1486.

Giampietro, V., Marquand, A., Brammer, M., Shawe-Taylor, J., & Mourao-Miranda, J. (2010). powerSVM: Towards power calculation in Machine Learning for brain imaging. Poster presented at the annual Human Brain Mapping Conference: Barcelona, Spain.

Gift, T. E., Strauss, J. S., Harder, D. W., Kokes, R. F., & Ritzler, B. A. (1981). Established chronicity of psychotic symptoms in first-admission schizophrenic patients. *The American journal of psychiatry, 138*(6), 779-784.

Giuliani, N. R., Calhoun, V. D., Pearlson, G. D., Francis, A., & Buchanan, R. W. (2005). Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophrenia research, 74*(2-3), 135-147.

Gong, Q., Wu, Q., Scarpazza, C., Lui, S., Jia, Z., Marquand, A., et al. (2011). Prognostic prediction of therapeutic response in depression using high-field MR imaging. *NeuroImage, 55*(4), 1497-1503.

Gore, F. M., Bloem, P. J. N., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., et al. (2011). Global burden of disease in young people aged 10-24 years: a systematic analysis. *Lancet, 377*(9783), 2093-2102.

Graña, M., Termenon, M., Savio, A., Gonzalez-Pinto, A., Echeveste, J., Pérez, J. M., et al. (2011). Computer aided diagnosis system for alzheimer disease using brain diffusion tensor imaging features selected by Pearson's correlation. *Neurosci Lett, 502*(3), 225-229.

Greenwood, T. A., Lazzeroni, L. C., Murray, S. S., Cadenhead, K. S., Calkins, M. E., Dobie, D. J., et al. (2011). Analysis of 94 candidate genes and 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. *The American journal of psychiatry, 168*(9), 930-946.

Groom, M. J., Jackson, G. M., Calton, T. G., Andrews, H. K., Bates, A. T., Liddle, P. F., et al. (2008). Cognitive deficits in early-onset schizophrenia spectrum patients and their non-psychotic siblings: a comparison with ADHD. *Schizophrenia research, 99*(1-3), 85-95.

Grosselin, A., Royer, A., Schneider, F. C. G., Brouillet, D., Martin, S., Pellet, J., et al. (2010). [Neuro-anatomic activations of prepotent responses in schizophrenia in Hayling's task]. *Encephale, 36*(4), 277-284.

Hall, J., Whalley, H. C., Job, D. E., Baig, B. J., McIntosh, A. M., Evans, K. L., et al. (2006). A neuregulin 1 variant associated with abnormal cortical function and psychotic symptoms. *Nat Neurosci, 9*(12), 1477-1478.

Hansen, T., Ingason, A., Djurovic, S., Melle, I., Fenger, M., Gustafsson, O., et al. (2011). At-risk variant in TCF7L2 for type II diabetes increases risk of schizophrenia. *Biol Psychiatry, 70*(1), 59-63.

Hastie, T., Tibshirani, R., Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* New York: NY: Springer-Verlag.

Hawkins, K. A., Addington, J., Keefe, R. S. E., Christensen, B., Perkins, D. O., Zipurksy, R., et al. (2004). Neuropsychological status of subjects at high risk for a first episode of psychosis. *Schizophr Res, 67*(2-3), 115-122.

Hawkins, K. A., Keefe, R. S. E., Christensen, B. K., Addington, J., Woods, S. W., Callahan, J., et al. (2008). Neuropsychological course in the prodrome and first episode of psychosis: findings from the PRIME North America Double Blind Treatment Study. *Schizophr Res, 105*(1-3), 1-9.

Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology, 12*(3), 426-445.

Hennah, W., Varilo, T., Kestilä, M., Paunio, T., Arajärvi, R., Haukka, J., et al. (2003). Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Hum Mol Genet, 12*(23), 3151-3159.

Herold, R., Feldmann, A., Simon, M., Tényi, T., Kövér, F., Nagy, F., et al. (2009). Regional gray matter reduction and theory of mind deficit in the early phase of schizophrenia: a voxel-based morphometric study. *Acta Psychiatr Scand, 119*(3), 199-208.

Hill, S. K., Beers, S. R., Kmiec, J. A., Keshavan, M. S., & Sweeney, J. A. (2004). Impairment of verbal memory and learning in antipsychotic-naïve patients with first-episode schizophrenia. *Schizophr Res, 68*(2-3), 127-136.

Hinrichs, C., Singh, V., Xu, G., & Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *Neuroimage, 55*(2), 574-589.

Honea, R., Crow, T. J., Passingham, D., & Mackay, C. E. (2005). Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry, 162*(12), 2233-2245.

Ingalhalikar, M., Kanterakis, S., Gur, R., Roberts, T. P. L., & Verma, R. (2010). DTI based diagnostic prediction of a disease via pattern classification. *Med Image Comput Comput Assist Interv, 13*(Pt 1), 558-565.

Iwashiro, N., Suga, M., Takano, Y., Inoue, H., Natsubori, T., Satomura, Y., et al. (2012). Localized gray matter volume reductions in the pars triangularis of the inferior frontal gyrus in individuals at clinical high-risk for psychosis and first episode for schizophrenia. *Schizophrenia research*.

Jablensky, A. (2000). Epidemiology of schizophrenia: the global burden of disease and disability. *European Archives of Psychiatry and Clinical …*.

James, A. C. D., Javaloyes, A., James, S., & Smith, D. M. (2002). Evidence for non-progressive changes in adolescent-onset schizophrenia: follow-up magnetic resonance imaging study. *Br J Psychiatry, 180*, 339-344.

Jastak, S., & Wilkinson, G. S. (1984). WRAT-R: Wide range achievement test administration manual. Los Angeles Western Psychological Services.

Job, D. E., Whalley, H. C., McConnell, S., Glabus, M., Johnstone, E. C., & Lawrie, S. M. (2002). Structural gray matter differences between first-episode schizophrenics and normal controls using voxel-based morphometry. *Neuroimage, 17*(2), 880-889.

Johnstone, B., Callahan, C. D., Kapila, C. J., & Bouman, D. E. (1996). The comparability of the WRAT-R reading test and NAART as estimates of premorbid intelligence in neurologically impaired patients. *Arch Clin Neuropsychol, 11*(6), 513-519.

Johnstone, E. C., Crow, T. J., Frith, C. D., Husband, J., & Kreel, L. (1976). Cerebral ventricular size and cognitive impairment in chronic schizophrenia. *Lancet, 2*(7992), 924-926.

Jones, D. K. (2008). Studying connections in the living human brain with diffusion MRI. *Cortex; a journal devoted to the study of the nervous system and behavior, 44*(8), 936-952.

Jones, D. K., Griffin, L. D., Alexander, D. C., Catani, M., Horsfield, M. A., Howard, R., et al. (2002). Spatial normalization and averaging of diffusion tensor MRI data sets. *Neuroimage, 17*(2), 592-617.

Jones, D. K., Simmons, A., Williams, S. C., & Horsfield, M. A. (1999). Non-invasive assessment of axonal fiber connectivity in the human brain via diffusion tensor MRI. *Magn Reson Med, 42*(1), 37-41.

Joshua, N., Gogos, A., & Rossell, S. (2009). Executive functioning in schizophrenia: a thorough examination of performance on the Hayling Sentence Completion Test compared to psychiatric and non-psychiatric controls. *Schizophrenia research, 114*(1-3), 84-90.

Jung, W. H., Kim, J. S., Jang, J. H., Choi, J.-S., Jung, M. H., Park, J.-Y., et al. (2011). Cortical thickness reduction in individuals at ultra-high-risk for psychosis. *Schizophr Bull, 37*(4), 839-849.

Karlsgodt, K. H., Niendam, T. A., Bearden, C. E., & Cannon, T. D. (2009). White matter integrity and prediction of social and role functioning in subjects at ultra-high risk for psychosis. *Biol Psychiatry, 66*(6), 562-569.

Kasparek, T., Prikryl, R., Schwarz, D., Kucerova, H., Marecek, R., Mikl, M., et al. (2009). Gray matter morphology and the level of functioning in one-year follow-up of first-episode schizophrenia patients. *Prog Neuropsychopharmacol Biol Psychiatry, 33*(8), 1438-1446.

Kawada, R., Yoshizumi, M., Hirao, K., Fujiwara, H., Miyata, J., Shimizu, M., et al. (2009). Brain volume and dysexecutive behavior in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry, 33*(7), 1255-1260.

Kay, S., Opler, L., Fiszbein, A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophr Bull*.

Keefe, R. S. E., Perkins, D. O., Gu, H., Zipursky, R. B., Christensen, B. K., & Lieberman, J. A. (2006). A longitudinal study of neurocognitive function in individuals at-risk for psychosis. *Schizophr Res, 88*(1-3), 26-35.

Kéri, S., Kiss, I., & Kelemen, O. (2009). Effects of a neuregulin 1 variant on conversion to schizophrenia and schizophreniform disorder in people at high risk for psychosis. *Mol Psychiatry, 14*(2), 118-119.

Keshavan, M. S., & Schooler, N. R. (1992). First-episode studies in schizophrenia: criteria and characterization. *Schizophrenia bulletin, 18*(3), 491-513.

Khodayari-Rostamabad, A., Hasey, G. M., Maccrimmon, D. J., Reilly, J. P., & de Bruin, H. (2010). A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. *Clin Neurophysiol, 121*(12), 1998-2006.

Kikinis, Z., Fallon, J. H., Niznikiewicz, M., Nestor, P., Davidson, C., Bobrow, L., et al. (2010). Gray matter volume reduction in rostral middle frontal gyrus in patients with chronic schizophrenia. *Schizophrenia research, 123*(2-3), 153-159.

Kindermann, S. S., Karimi, A., Symonds, L., Brown, G. G., & Jeste, D. V. (1997). Review of functional magnetic resonance imaging in schizophrenia. *Schizophrenia research, 27*(2-3), 143-156.

Kircher, T. T., Bulimore, E. T., Brammer, M. J., Williams, S. C., Broome, M. R., Murray, R. M., et al. (2001). Differential activation of temporal cortex during sentence completion in schizophrenic patients with and without formal thought disorder. *Schizophr Res, 50*(1-2), 27-40.

Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers (Vol. 20, pp. 226-239): Pattern Analysis and Machine Intelligence, IEEE Transactions on.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science, 308*(5720), 385-389.

Klosterkötter, J., Hellmich, M., Steinmeyer, E. M., & Schultze-Lutter, F. (2001). Diagnosing schizophrenia in the initial prodromal phase. *Arch Gen Psychiatry, 58*(2), 158-164.

Knöchel, C., Oertel-Knöchel, V., Schönmeyer, R., Rotarska-Jagiela, A., van de Ven, V., Prvulovic, D., et al. (2012). Interhemispheric hypoconnectivity in schizophrenia: fiber integrity and volume differences of the corpus callosum in patients and unaffected relatives. *NeuroImage, 59*(2), 926-934.

Koutsouleris, N., Borgwardt, S., Meisenzahl, E. M., Bottlender, R., Möller, H.-J., & Riecher-Rössler, A. (2011a). Disease Prediction in the At-Risk Mental State for Psychosis Using Neuroanatomical Biomarkers: Results From the FePsy Study. *Schizophrenia bulletin*.

Koutsouleris, N., Davatzikos, C., Bottlender, R., Patschurek-Kliche, K., Scheuerecker, J., Decker, P., et al. (2011b). Early Recognition and Disease Prediction in the At-Risk Mental States for Psychosis Using Neurocognitive Pattern Classification. *Schizophr Bull*.

Koutsouleris, N., Meisenzahl, E. M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., et al. (2009a). Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry, 66*(7), 700-712.

Koutsouleris, N., Patschurek-Kliche, K., Scheuerecker, J., Decker, P., Bottlender, R., Schmitt, G., et al. (2010). Neuroanatomical correlates of executive dysfunction in the at-risk mental state for psychosis. *Schizophrenia research, 123*(2-3), 160-174.

Koutsouleris, N., Schmitt, G. J. E., Gaser, C., Bottlender, R., Scheuerecker, J., McGuire, P., et al. (2009b). Neuroanatomical correlates of different vulnerability states for psychosis and their clinical outcomes. *Br J Psychiatry, 195*(3), 218-226.

Krug, A., Nieratschker, V., Markov, V., Krach, S., Jansen, A., Zerres, K., et al. (2010). Effect of CACNA1C rs1006737 on neural correlates of verbal fluency in healthy individuals. *Neuroimage, 49*(2), 1831-1836.

Kubicki, M., Shenton, M. E., Salisbury, D. F., Hirayasu, Y., Kasai, K., Kikinis, R., et al. (2002). Voxel-based morphometric analysis of gray matter in first episode schizophrenia. *Neuroimage, 17*(4), 1711-1719.

Kunimatsu, N., Aoki, S., Kunimatsu, A., Abe, O., Yamada, H., Masutani, Y., et al. (2012). Tract-specific analysis of white matter integrity disruption in schizophrenia. *Psychiatry Res, 201*(2), 136-143.

Lanckriet, G., Cristianini, N., Bartlett, P., Ghaoui, L., & Jordan, M. (2004). Learning the Kernel Matrix with Semidefinite Programming (Vol. 5, pp. 27-72): Journal of Machine Learning Research.

Lao, Z., Shen, D., Xue, Z., Karacali, B., & Resnick, S. (2004). Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*.

Leemans, A., Jeurissen, B., Sijbers, J., & Jones, D. (2009). ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data (pp. p. 3537). Hawaii, USA: In: 17th Annual Meeting of Intl Soc Mag Reson Med.

Leemans, A., & Jones, D. K. (2009). The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med, 61*(6), 1336-1349.

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *Neuroimage, 56*(2), 387-399.

Lencz, T., Smith, C. W., McLaughlin, D., Auther, A., Nakayama, E., Hovey, L., et al. (2006). Generalized and specific neurocognitive deficits in prodromal schizophrenia. *Biol Psychiatry, 59*(9), 863-871.

Lewis, C. M., & Knight, J. (2012). Introduction to genetic association studies. *Cold Spring Harb Protoc, 2012*(3), 297-306.

Lewis, D. P., Jebara, T., & Noble, W. S. (2006). Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure. *Bioinformatics, 22*(22), 2753-2760.

Li, D., Collier, D. A., & He, L. (2006). Meta-analysis shows strong positive association of the neuregulin 1 (NRG1) gene with schizophrenia. *Hum Mol Genet, 15*(12), 1995-2002.

Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magn Reson Imaging, 22*(10), 1517-1531.

Luck, D., Buchy, L., Czechowska, Y., Bodnar, M., Pike, G. B., Campbell, J. S. W., et al. (2010). Fronto-temporal disconnectivity and clinical short-term outcome in first episode psychosis: A DTI-tractography study. *Journal of psychiatric research*.

Lunetta, K. L. (2008). Genetic association studies. *Circulation, 118*(1), 96-101.

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., et al. (2009). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology, 51*(2), 73-83.

Mansfield, P. (1984). Real-time echo-planar imaging by NMR. *Br Med Bull, 40*(2), 187-190.

Marcezewski, P., Vanr de Linden, M., & Larøi, F. (2001). Further investigation of the Supervisory Attentional System in schizophrenia: Planning, inhibition, and rule abstraction (pp. 175-192): Cognitive Neuropsychiatry.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., & Mourão-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *NeuroImage, 49*(3), 2178-2189.

Marquand, A. F., Mourão-Miranda, J., Brammer, M. J., Cleare, A. J., & Fu, C. H. Y. (2008). Neuroanatomy of verbal working memory as a diagnostic biomarker for depression. *Neuroreport, 19*(15), 1507-1511.

Mathalon, D. H., Sullivan, E. V., Lim, K. O., & Pfefferbaum, A. (2001). Progressive brain volume changes and the clinical course of schizophrenia in men: a longitudinal magnetic resonance imaging study. *Arch Gen Psychiatry, 58*(2), 148-157.

Matthews, P. M., Honey, G. D., & Bullmore, E. T. (2006). Applications of fMRI in translational medicine and clinical practice. *Nat Rev Neurosci, 7*(9), 732-744.

McGrath, J., Saha, S., & Chant, D. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiologic reviews*.

McGuffin, P., Owen, M. J., & Farmer, A. E. (1995). Genetic basis of schizophrenia. *Lancet, 346*(8976), 678-682.

Mechelli, A., Riecher-Rössler, A., Meisenzahl, E. M., Tognin, S., Wood, S. J., Borgwardt, S. J., et al. (2011). Neuroanatomical abnormalities that predate the onset of psychosis: a multicenter study. *Arch Gen Psychiatry, 68*(5), 489-495.

Meisenzahl, E. M., Koutsouleris, N., Bottlender, R., Scheuerecker, J., Jäger, M., Teipel, S. J., et al. (2008a). Structural brain alterations at different stages of schizophrenia: a voxel-based morphometric study. *Schizophrenia research, 104*(1-3), 44-60.

Meisenzahl, E. M., Koutsouleris, N., Gaser, C., Bottlender, R., Schmitt, G. J. E., McGuire, P., et al. (2008b). Structural brain alterations in subjects at high-risk of psychosis: a voxel-based morphometric study. *Schizophrenia research, 102*(1-3), 150-162.

Menon, R. R., Barta, P. E., Aylward, E. H., Richards, S. S., Vaughn, D. D., Tien, A. Y., et al. (1995). Posterior superior temporal gyrus in schizophrenia: grey matter changes and clinical correlates. *Schizophrenia research, 16*(2), 127-135.

Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V., & Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: a meta-analytic review. *Neuropsychology, 23*(3), 315-336.

Miller, T. J., McGlashan, T. H., Rosen, J. L., Somjee, L., Markovich, P. J., Stein, K., et al. (2002). Prospective diagnosis of the initial prodrome for schizophrenia based on the Structured Interview for Prodromal Syndromes: preliminary evidence of interrater reliability and predictive validity. *The American journal of psychiatry, 159*(5), 863-865.

Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S., & Glahn, D. C. (2009). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Arch Gen Psychiatry, 66*(8), 811-822.

Molina, V., Sanz, J., Villa, R., Pérez, J., González, D., Sarramea, F., et al. (2010). Voxel-based morphometry comparison between first episodes of psychosis with and without evolution to schizophrenia. *Psychiatry Res, 181*(3), 204-210.

Morey, R. A., Inan, S., Mitchell, T. V., Perkins, D. O., Lieberman, J. A., & Belger, A. (2005). Imaging frontostriatal function in ultra-high-risk, early, and chronic schizophrenia during executive processing. *Arch Gen Psychiatry, 62*(3), 254-262.

Morgan, K. D., Dazzan, P., Orr, K. G., Hutchinson, G., Chitnis, X., Suckling, J., et al. (2007). Grey matter abnormalities in first-episode schizophrenia and affective psychosis. *Br J Psychiatry Suppl, 51*, s111-116.

Moseley, M. E., Cohen, Y., Kucharczyk, J., Mintorovitch, J., Asgari, H. S., Wendland, M. F., et al. (1990). Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology, 176*(2), 439-445.

Mourão-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *NeuroImage, 28*(4), 980-995.

Nathaniel-James, D. A., Fletcher, P., Frith, C.D. (1997). The functional anatomy of verbal initiation and suppression using the Hayling Test. *Neuropsychologia*, 559-566.

Navari, S., & Dazzan, P. (2009). Do antipsychotic drugs affect brain structure? A systematic and critical review of MRI findings. *Psychological medicine*.

Neckelmann, G., Specht, K., Lund, A., Ersland, L., Smievoll, A. I., Neckelmann, D., et al. (2006). Mr morphometry analysis of grey matter volume reduction in schizophrenia: association with hallucinations. *Int J Neurosci, 116*(1), 9-23.

Nho, K., Shen, L., Kim, S., Risacher, S. L., West, J. D., Foroud, T., et al. (2010). Automatic Prediction of Conversion from Mild Cognitive Impairment to Probable Alzheimer's Disease using Structural Magnetic Resonance Imaging. *AMIA Annu Symp Proc, 2010*, 542-546.

Niendam, T. A., Bearden, C. E., Zinberg, J., Johnson, J. K., O'Brien, M., & Cannon, T. D. (2007). The course of neurocognition and social functioning in individuals at ultra high risk for psychosis. *Schizophr Bull, 33*(3), 772-781.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci (Regul Ed), 10*(9), 424-430.

O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., et al. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet, 40*(9), 1053-1055.

Ogawa, S., & Lee, T. M. (1990). Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. *Magn Reson Med, 16*(1), 9-18.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990a). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci USA, 87*(24), 9868-9872.

Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990b). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med, 14*(1), 68-78.

Ogawa, S., Menon, R. S., Tank, D. W., Kim, S. G., Merkle, H., Ellermann, J. M., et al. (1993). Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys J, 64*(3), 803-812.

Oliveira, P. P. d. M., Nitrini, R., Busatto, G., Buchpiguel, C., Sato, J. R., & Amaro, E. (2010). Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease. *J Alzheimers Dis, 19*(4), 1263-1272.

Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience and biobehavioral reviews, 36*(4), 1140-1152.

Pantelis, C., Velakoulis, D., McGorry, P. D., Wood, S. J., Suckling, J., Phillips, L. J., et al. (2003). Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison. *Lancet, 361*(9354), 281-288.

Pauli, A., Prata, D. P., Mechelli, A., Picchioni, M., Fu, C. H. Y., Chaddock, C. A., et al. (2012). Interaction between effects of genes coding for dopamine and glutamate transmission on striatal and parahippocampal function. *Human brain mapping*.

Paulsen, J. S., Heaton, R. K., Sadek, J. R., Perry, W., Delis, D. C., Braff, D., et al. (1995). The nature of learning and memory impairments in schizophrenia. *J Int Neuropsychol Soc, 1*(1), 88-99.

Peers, I. (1996). *Statistical Analysis for Education and Psychology Researchers*: Falmer Press.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage, 45*(1 Suppl), S199-209.

Peters, B. D., de Haan, L., Dekker, N., Blaas, J., Becker, H. E., Dingemans, P. M., et al. (2008). White matter fibertracking in first-episode schizophrenia, schizoaffective patients and subjects at ultra-high risk of psychosis. *Neuropsychobiology, 58*(1), 19-28.

Peters, B. D., Dingemans, P. M., Dekker, N., Blaas, J., Akkerman, E., van Amelsvoort, T. A., et al. (2010). White matter connectivity and psychosis in ultra-high-risk subjects: A diffusion tensor fiber tracking study. *Psychiatry Res, 181*(1), 44-50.

Peters, B. D., Schmitz, N., Dingemans, P. M., van Amelsvoort, T. A., Linszen, D. H., de Haan, L., et al. (2009). Preliminary evidence for reduced frontal white matter integrity in subjects at ultra-high-risk for psychosis. *Schizophrenia research, 111*(1-3), 192-193.

Pettersson-Yeo, W., Allen, P., Benetti, S., McGuire, P., & Mechelli, A. (2011). Dysconnectivity in schizophrenia: where are we now? *Neuroscience and biobehavioral reviews, 35*(5), 1110-1124.

Phillips, C. L., Bruno, M.-A., Maquet, P., Boly, M., Noirhomme, Q., Schnakers, C., et al. (2011). "Relevance vector machine" consciousness classifier applied to cerebral metabolism of vegetative and locked-in patients. *NeuroImage, 56*(2), 797-808.

Phillips, L. J., Yung, A. R., & McGorry, P. D. (2000). Identification of young people at risk of psychosis: validation of Personal Assessment and Crisis Evaluation Clinic intake criteria. *Aust N Z J Psychiatry, 34 Suppl*, S164-169.

Pierpaoli, C., Jezzard, P., Basser, P. J., Barnett, A., & Di Chiro, G. (1996). Diffusion tensor MR imaging of the human brain. *Radiology, 201*(3), 637-648.

Plant, C., Teipel, S. J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., et al. (2010). Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *NeuroImage, 50*(1), 162-174.

Pomarol-Clotet, E., Canales-Rodríguez, E. J., Salvador, R., Sarró, S., Gomar, J. J., Vila, F., et al. (2010). Medial prefrontal cortex pathology in schizophrenia as revealed by convergent findings from multimodal imaging. *Mol Psychiatry, 15*(8), 823-830.

Prata, D. P., Mechelli, A., Fu, C. H. Y., Picchioni, M., Toulopoulou, T., Bramon, E., et al. (2009). Epistasis between the DAT 3' UTR VNTR and the COMT Val158Met

SNP on cortical function in healthy subjects and patients with schizophrenia. *Proc Natl Acad Sci USA, 106*(32), 13600-13605.

Price, G., Bagary, M. S., Cercignani, M., Altmann, D. R., & Ron, M. A. (2005). The corpus callosum in first episode schizophrenia: a diffusion tensor imaging study. *J Neurol Neurosurg Psychiatr, 76*(4), 585-587.

Price, G., Cercignani, M., Chu, E. M., Barnes, T. R. E., Barker, G. J., Joyce, E. M., et al. (2010). Brain pathology in first-episode psychosis: magnetization transfer imaging provides additional information to MRI measurements of volume loss. *Neuroimage, 49*(1), 185-192.

Price, G., Cercignani, M., Parker, G. J. M., Altmann, D. R., Barnes, T. R. E., Barker, G. J., et al. (2008). White matter tracts in first-episode psychosis: a DTI tractography study of the uncinate fasciculus. *NeuroImage, 39*(3), 949-955.

Puddephat, M. (2010). http://www.mikepuddephat.com/Page/1603/principles-of-magnetic-resonance-imaging.

Pukrop, R., Ruhrmann, S., Schultze-Lutter, F., Bechdolf, A., Brockhaus-Dumke, A., & Klosterkötter, J. (2007). Neurocognitive indicators for a conversion to psychosis: comparison of patients in a potentially initial prodromal state who did or did not convert to a psychosis. *Schizophrenia research, 92*(1-3), 116-125.

Rakotomamonjy, A., Bach, F., Canu, S., & Grandvalet, Y. (2008). Simple MKL (Vol. 9, pp. 2491-2521). J. Mach. Learn. Res.

Rasetti, R., Sambataro, F., Chen, Q., Callicott, J. H., Mattay, V. S., & Weinberger, D. R. (2011). Altered cortical network dynamics: a potential intermediate phenotype for schizophrenia and association with ZNF804A. *Arch Gen Psychiatry, 68*(12), 1207-1217.

Rasmussen, C., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning: The MIT Press, Cambridge, MA.

Riecher-Rössler, A., Aston, J., Ventura, J., Merlo, M., Borgwardt, S., Gschwandtner, U., et al. (2008). [The Basel Screening Instrument for Psychosis (BSIP): development, structure, reliability and validity]. *Fortschr Neurol Psychiatr, 76*(4), 207-216.

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat Genet, 43*(10), 969-976.

Rivero, O., Reif, A., Sanjuán, J., Moltó, M. D., Kittel-Schneider, S., Nájera, C., et al. (2010). Impact of the AHI1 gene on the vulnerability to schizophrenia: a case-control association study. *PLoS ONE, 5*(8), e12254.

Ruhrmann, S., Schultze-Lutter, F., Maier, W., & Klosterkötter, J. (2005). Pharmacological intervention in the initial prodromal phase of psychosis. *Eur Psychiatry, 20*(1), 1-6.

Salgado-Pineda, P., Baeza, I., Pérez-Gómez, M., Vendrell, P., Junqué, C., Bargalló, N., et al. (2003). Sustained attention impairment correlates to gray matter decreases in first episode neuroleptic-naive schizophrenic patients. *Neuroimage, 19*(2 Pt 1), 365-375.

Sartori, G., Pellegrini, S., & Mechelli, A. (2011). Forensic neurosciences: from basic research to applications and pitfalls. *Curr Opin Neurol, 24*(4), 371-377.

Schneider, F. C., Royer, A., Grosselin, A., Pellet, J., Barral, F.-G., Laurent, B., et al. (2011). Modulation of the default mode network is task-dependant in chronic schizophrenia patients. *Schizophr Res, 125*(2-3), 110-117.

Schölkopf, B., & Smola, A. J. (2002). Learning with kernels: support vector machines, regularization, optimization, and beyond.

Schultze-Lutter, F. (2009). Subjective symptoms of schizophrenia in research and the clinic: the basic symptom concept. *Schizophr Bull, 35*(1), 5-8.

Schulze, T. G., Detera-Wadleigh, S. D., Akula, N., Gupta, A., Kassem, L., Steele, J., et al. (2009). Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. *Mol Psychiatry, 14*(5), 487-491.

Segal, D., Haznedar, M. M., Hazlett, E. A., Entis, J. J., Newmark, R. E., Torosjan, Y., et al. (2010). Diffusion tensor anisotropy in the cingulate gyrus in schizophrenia. *NeuroImage*.

Sei, Y., Li, Z., Song, J., Ren-Patterson, R., Tunbridge, E. M., Iizuka, Y., et al. (2010). Epistatic and functional interactions of catechol-o-methyltransferase (COMT) and AKT1 on neuregulin1-ErbB signaling in cell models. *PLoS ONE, 5*(5), e10789.

Seidman, L. J., Giuliano, A. J., Meyer, E. C., Addington, J., Cadenhead, K. S., Cannon, T. D., et al. (2010). Neuropsychology of the prodrome to psychosis in the NAPLS consortium: relationship to family history and conversion to psychosis. *Arch Gen Psychiatry, 67*(6), 578-588.

Shapleske, J., Rossell, S. L., Chitnis, X. A., Suckling, J., Simmons, A., Bullmore, E. T., et al. (2002). A computational morphometric MRI study of schizophrenia: effects of hallucinations. *Cereb Cortex, 12*(12), 1331-1341.

Smieskova, R., Fusar-Poli, P., Allen, P., Bendfeldt, K., Stieglitz, R. D., Drewe, J., et al. (2010). Neuroimaging predictors of transition to psychosis-A systematic review and meta-analysis. *Neuroscience and biobehavioral reviews*.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., et al. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage, 31*(4), 1487-1505.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage, 44*(1), 83-98.

Sonnenburg, S., Rätsch, G., Schäfer, C., & Schölkopf, B. (2006). Large Scale Multiple Kernel Learning, *J. Mach. Learn. Res.* (Vol. 7, pp. 1531-1565).

Sponheim, S. R., Jung, R. E., Seidman, L. J., Mesholam-Gately, R. I., Manoach, D. S., O'Leary, D. S., et al. (2010). Cognitive deficits in recent-onset and chronic schizophrenia. *J Psychiatr Res, 44*(7), 421-428.

Stabenau, J. R., & Pollin, W. (1967). Early characteristics of monozygotic twins discordant for schizophrenia. *Arch Gen Psychiatry, 17*(6), 723-734.

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., et al. (2009). Common variants conferring risk of schizophrenia. *Nature, 460*(7256), 744-747.

Steinberg, S., de Jong, S., Consortium, I. S. G., Andreassen, O. A., Werge, T., Børglum, A. D., et al. (2011). Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet, 20*(20), 4076-4081.

Stone, J. M., Day, F., Tsagaraki, H., Valli, I., McLean, M. A., Lythgoe, D. J., et al. (2009). Glutamate dysfunction in people with prodromal symptoms of psychosis: relationship to gray matter volume. *Biological psychiatry, 66*(6), 533-539.

Stonnington, C. M., Chu, C., Klöppel, S., Jack, C. R., Ashburner, J., Frackowiak, R. S. J., et al. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage, 51*(4), 1405-1413.

Sun, D., van Erp, T. G. M., Thompson, P. M., Bearden, C. E., Daley, M., Kushan, L., et al. (2009). Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. *Biological psychiatry, 66*(11), 1055-1060.

Tanskanen, P., Ridler, K., Murray, G. K., Haapea, M., Veijola, J. M., Jääskeläinen, E., et al. (2010). Morphometric brain abnormalities in schizophrenia in a population-based sample: relationship to duration of illness. *Schizophr Bull, 36*(4), 766-777.

Thames Valley Test Company Limited. (1997). The Hayling and Brixton Tests. London: Battley Brothers Limited.

Thiebaut de Schotten, M., ffytche, D. H., Bizzi, A., Dell'Acqua, F., Allin, M., Walshe, M., et al. (2011). Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *NeuroImage, 54*(1), 49-59.

Tipping, M. (2001). Sparse Bayesian learning and the relevance vector machine (Vol. 1, pp. 211-244). J. Mach. Learn. Res.

Tracy, J. I., Mattson, R., King, C., Bundick, T., Celenza, M. A., & Glosser, G. (2001). A comparison of memory for verbal and non-verbal material in schizophrenia. *Schizophr Res, 50*(3), 199-211.

Tyson, P. J., Laws, K. R., Roberts, K. H., & Mortimer, A. M. (2004). Stability of set-shifting and planning abilities in patients with schizophrenia. *Psychiatry research, 129*(3), 229-239.

van Haren, N. E. M., Hulshoff Pol, H. E., Schnack, H. G., Cahn, W., Brans, R., Carati, I., et al. (2008). Progressive brain volume loss in schizophrenia over the course of the illness: evidence of maturational abnormalities in early adulthood. *Biological psychiatry, 63*(1), 106-113.

van Haren, N. E. M., Rijsdijk, F., Schnack, H. G., Picchioni, M. M., Toulopoulou, T., Weisbrod, M., et al. (2012). The Genetic and Environmental Determinants of the Association Between Brain Abnormalities and Schizophrenia: The Schizophrenia Twins and Relatives Consortium. *Biological psychiatry*.

van Os, J., & Kapur, S. (2009). Schizophrenia. *Lancet, 374*(9690), 635-645.

Vanzetta, I., Slovin, H., Omer, D. B., & Grinvald, A. (2004). Columnar resolution of blood volume and oximetry functional maps in the behaving monkey; implications for FMRI. *Neuron, 42*(5), 843-854.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council, 10*(5), 988-999.

Velakoulis, D., Wood, S. J., Wong, M. T. H., McGorry, P. D., Yung, A., Phillips, L., et al. (2006). Hippocampal and amygdala volumes according to psychosis stage and diagnosis: a magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra-high-risk individuals. *Arch Gen Psychiatry, 63*(2), 139-149.

Walterfang, M., Yung, A., Wood, A. G., Reutens, D. C., Phillips, L., Wood, S. J., et al. (2008). Corpus callosum shape alterations in individuals prior to the onset of psychosis. *Schizophrenia research, 103*(1-3), 1-10.

Walton, E., Turner, J., Gollub, R. L., Manoach, D. S., Yendiki, A., Ho, B.-C., et al. (2012). Cumulative Genetic Risk and Prefrontal Activity in Patients With Schizophrenia. *Schizophrenia bulletin*.

Whalley, H. C., Simonotto, E., Flett, S., Marshall, I., Ebmeier, K. P., Owens, D. G. C., et al. (2004). fMRI correlates of state and trait effects in subjects at genetically enhanced risk of schizophrenia. *Brain, 127*(Pt 3), 478-490.

Whalley, H. C., Simonotto, E., Marshall, I., Owens, D. G. C., Goddard, N. H., Johnstone, E. C., et al. (2005). Functional disconnectivity in subjects at high genetic risk of schizophrenia. *Brain, 128*(Pt 9), 2097-2108.

Wilkinson, D., & Halligan, P. (2004). The relevance of behavioural measures for functional-imaging studies of cognition. *Nat Rev Neurosci, 5*(1), 67-73.

Williams, H. J., Craddock, N., Russo, G., Hamshere, M. L., Moskvina, V., Dwyer, S., et al. (2011). Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. *Hum Mol Genet, 20*(2), 387-391.

Williams, N. M., Preece, A., Morris, D. W., Spurlock, G., Bray, N. J., Stephens, M., et al. (2004). Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry, 61*(4), 336-344.

Witthaus, H., Brüne, M., Kaufmann, C., Bohner, G., Ozgürdal, S., Gudlowski, Y., et al. (2008). White matter abnormalities in subjects at ultra high-risk for schizophrenia and first-episode schizophrenic patients. *Schizophrenia research, 102*(1-3), 141-149.

Witthaus, H., Kaufmann, C., Bohner, G., Ozgürdal, S., Gudlowski, Y., Gallinat, J., et al. (2009). Gray matter abnormalities in subjects at ultra-high risk for schizophrenia and first-episode schizophrenic patients compared to healthy controls. *Psychiatry Res, 173*(3), 163-169.

Wood, S. J., Brewer, W. J., Koutsouradis, P., Phillips, L. J., Francey, S. M., Proffitt, T. M., et al. (2007a). Cognitive decline following psychosis onset: data from the PACE clinic. *The British journal of psychiatry Supplement, 51*, s52-57.

Wood, S. J., Tarnawski, A. U., Proffitt, T. M., Brewer, W. J., Savage, G. R., Anderson, V., et al. (2007b). Fractionation of verbal memory impairment in schizophrenia and schizophreniform psychosis. *Aust N Z J Psychiatry, 41*(9), 732-739.

Wood, S. J., Yücel, M., Velakoulis, D., Phillips, L. J., Yung, A. R., Brewer, W., et al. (2005). Hippocampal and anterior cingulate morphology in subjects at ultra-high-risk for psychosis: the role of family history of psychotic illness. *Schizophrenia research, 75*(2-3), 295-301.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping, 4*(1), 58-73.

Yang, H., Liu, J., Sui, J., Pearlson, G., & Calhoun, V. D. (2010). A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining both Improves Classification of Schizophrenia. *Front Hum Neurosci, 4*, 192.

Yassa, M. A., & Stark, C. E. L. (2009). A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *Neuroimage, 44*(2), 319-327.

Yung, A. R., Nelson, B., Stanford, C., Simmons, M. B., Cosgrave, E. M., Killackey, E., et al. (2008). Validation of "prodromal" criteria to detect individuals at ultra high risk of psychosis: 2 year follow-up. *Schizophrenia research, 105*(1-3), 10-17.

Yung, A. R., Phillips, L. J., McGorry, P. D., McFarlane, C. A., Francey, S., Harrigan, S., et al. (1998). Prediction of psychosis. A step towards indicated prevention of schizophrenia, *The British journal of psychiatry Supplement* (Vol. 172, pp. 14-20).

Yung, A. R., Phillips, L. J., Yuen, H. P., Francey, S. M., McFarlane, C. A., Hallgren, M., et al. (2003). Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group. *Schizophrenia research, 60*(1), 21-32.

Zakzanis, K. K., Graham, S. J., & Campbell, Z. (2003). A meta-analysis of structural and functional brain imaging in dementia of the Alzheimer's type: a neuroimaging profile. *Neuropsychol Rev, 13*(1), 1-18.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., & Initiative, A. s. D. N. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage, 55*(3), 856-867.

Zimmerman, D. (1997). A Note on Interpretation of the Paired-Samples t Test. *Journal of Educational and Behavioral Statistics, 22*(3), 349-360.

Zinkstok, J., Schmitz, N., van Amelsvoort, T., Moeton, M., Baas, F., & Linszen, D. (2008). Genetic variation in COMT and PRODH is associated with brain anatomy in patients with schizophrenia. *Genes Brain Behav, 7*(1), 61-69.

APPENDIX

A: Publications Arising

"Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high risk and first episode psychosis at individual level", Pettersson-Yeo W., Benetti S., Marquand A.F., Dell'Acqua F., Williams S.C-R, Allen P., Prata D., McGuire P., Mechelli A., *Psychological Medicine* (Accepted for Publication)

"An empirical comparison of different approaches for combining multimodal neuroimaging data with Support Vector Machine", Pettersson-Yeo W., Benetti S., Marquand A.F., Catani M., Williams S.C-R, Allen P., McGuire P., Mechelli A., (In Preparation)